University of Alberta

# The Perceived Role of Diploma Examinations in Alberta High Schools

by

Marlene Anne McDonald

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of Master of Education
in Educational Administration and Leadership

Department of Educational Policy Studies

Edmonton, Alberta

Spring 2001

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-60397-0

Canada

University of Alberta

Library Release Form

Name of Author: Marlene Anne McDonald

Title of Thesis: The Perceived Role of Diploma Examinations in Alberta High Schools

Degree: Master of Education

Year this Degree Granted: 2001

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

6242-112 A Street
Edmonton, AB
T6H 3K3

April 17, 2001

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommended to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled The Perceived Role of Diploma Examinations in Alberta High Schools submitted by Marlene Anne McDonald in partial fulfillment of the requirements for the degree of Master of Education in Educational Administration and Leadership.

M Haughey

Margaret Haughey, PhD
Supervisor

G. D. Taylor

Gerald Taylor, PhD

Carolyn Yewchuk

Carolyn Yewchuk, PhD

April 11, 2001

ABSTRACT

The purpose of this study was to investigate the perceived role of diploma examination results by Alberta high school teachers and to replicate and extend Loerke's 1993 study. Classroom teachers and school-based administrators (n=452) across Alberta took part in the research questionnaire developed to measure the variables under study. Findings from this study indicate that: classroom instruction has improved as a result of the re-implementation of diploma examinations, concerns about teacher autonomy persist, study participants lacked the professional training necessary to make the most of the available student assessment information, and the perceived role of diploma examination results has changed since the 1993 Loerke study. These findings have the potential to contribute to changes in teacher assessment practices in Alberta with respect to the appropriate use of standardized test results as a measure of student learning.

TABLE OF CONTENTS


CHAPTER 1 - OVERVIEW AND PURPOSE OF THE STUDY....1

PURPOSE OF THE STUDY....4

SIGNIFICANCE OF THE STUDY....5

SCOPE OF THE STUDY....6

CHAPTER 2 - LITERATURE REVIEW....8

REFORM BY COMPARISON....8

Introduction of Large-Scale Testing....13

The Politics of Accountability....15

Impact of Testing on Teaching Practices....17

Diploma Re-implementation in Alberta....19

IMPACT OF MEASUREMENT-DRIVEN INSTRUCTION....21

Instruction Diverted....21

Effect on the Behavior of Students, Teachers,

and Administrators....26

Setting Targets....27

Uses of Test Scores....29

Student Results and Ranking....30

Media Impact....34

Test Quality....36

IMPROVING TEST SCORES....40

Need for Professional Development....43

Cashing in on Good Scores....49

AN ACCOUNTABILITY-DRIVEN REFORM....50

Teacher Compensation....51
Personnel Practices Tied to Student Achievement....52
Alberta Incentive Proposal....53
Reactions to SPIP....54
A Counter Proposal....56
SPIP– A Closer Look....58
Rewards and Punishment....61
SUMMARY....64
1993 DIPLOMA EXAMINATION STUDY....65
CHAPTER 3 - RESEARCH DESIGN AND METHOD....68
RESEARCH DESIGN....68
LOERKE (1993) STUDY....68
SELECTION OF PARTICIPANTS....71
Rate of Return....72
Profile of Respondents....72
SELECTION AND PROFILE OF INTERVIEWEES....73
DATA COLLECTION....73
Pilot Study....74
Questionnaire....74
Interviews....75
Timeline....76
PROTECTION OF HUMAN SUBJECTS....77
DATA ANALYSIS....80
Factor Solution Analysis....81
Qualitative Data Analysis....82

VALIDITY AND RELIABILITY....83
SUMMARY....84
CHAPTER 4 - FINDINGS....86
RESEARCH QUESTIONS....89
PERCEIVED ROLE OF DIPLOMA EXAMINATIONS....94
Factor 1- Staff Evaluation....94
Factor 2- Impact on Instruction....100
Factor 3- Course Selection....102
Factor 4- Assessment Expertise....103
Factor 5- Teacher Stress....104
Factor 6- Staff Relations....105
Factor 7- Results Recognition....106
Factor 8- Improving Results....107
Factor 9- Use of Results....108
Qualitative Findings....108
PERCEPTIONS OF VARIOUS EDUCATIONAL GROUPS....110
Factors by School Position....111
Factors by School Size....113
Factors by Courses Taught....115
Factors by School Role....117
Summary of Group Perceptions....119
HOW WELL INFORMED ARE ALBERTA TEACHERS?....129
Awareness of Alberta Learning Documents....129
Assessment Training....130

Results Sharing....131
School/District Policy....132
Effect of the Media....134
Understanding Demographics....135
Ranking....135
School Selection....136
STUDENT RESULTS AS A MEASURE OF TEACHER....
COMPETENCY....136
STUDY COMPARISON....142
1993 Research Results....143
Loerke(1993 ) Study Comparison....145
Summary of the Study Comparison....157
SUMMARY....158
CHAPTER 5 -SUMMARY, CONCLUSIONS, AND IMPLICATIONS....162
OVERVIEW....162
PURPOSE OF THE STUDY....162
SUMMARY OF THE FINDINGS....163
Perceived Role of Diploma Examination Results....163
Perceived Differences Among Respondent Categories....168
Assessment Expertise of Alberta Educators....169
Use of Results in Teacher Assessment....171
Loerke (1993) Study Comparison....173
DISCUSSION OF THE FINDINGS....175
Perceived Role of Diploma Examination Results....175

Loerke (1993) Study Comparison....180
CONCLUSIONS....183
IMPLICATIONS FOR RESEARCH AND PRACTICE....183
Implications for Further Research....183
Implications for Practice....185
CONCLUDING COMMENT....188
REFERENCES....190
APPENDIX A - QUESTIONNAIRE....202
Questionnaire....203
APPENDIX B - STUDY CORRESPONDENCE....209
Letter to Superintendents....210
Letter to Principals....212
Letter to Participants....214
APPENDIX C - INTERVIEWS AND CORRESPONDENCE....215
Interview Participation Letter....216
Interview Format....218
Interview Questions....218
APPENDIX D - STATISTICAL TABLES....221
Appendix 1 Individual Item Analysis....222
Appendix 2 Comparison of Percentage Distribution....229
Appendix 3 Common Study Items....232
Appendix 4 Total Variance Explained for Factors....233
Appendix 5 Study Comparison....234

LIST OF TABLES


TABLE 1 Varimax Factor Solution for the Role of Diploma Examinations Using 9 Factors....89

TABLE 2 Varimax Factor Solution for the Role of Diploma Examinations Using 9 Factors, Means and Standard Deviations....95

TABLE 3 T-tests of Factor Scores of Measures Acceptable in Principle for the Role of Diploma Examinations by School Position....112

TABLE 4 One-way Analysis of Variance of Factor Scores of Measures Acceptable in Principle for the Role of Diploma Examinations by School Size....114

TABLE 5 T-tests of Factor Scores of Measures Acceptable in Principle for the Role of Diploma Examinations by Teacher Category....116

TABLE 6 One-way Analysis of Variance of Factor Scores of Measures Acceptable in Principle for the Role of Diploma Examinations by School Role....118

LIST OF FIGURES


| Figure | | Page |
|---|---|---|
| 1 | Means by Reporting Category for Factor 1 | 120 |
| 2 | Means by Reporting Category for Factor 2 | 121 |
| 3 | Means by Reporting Category for Factor 3 | 122 |
| 4 | Means by Reporting Category for Factor 4 | 123 |
| 5 | Means by Reporting Category for Factor 5 | 124 |
| 6 | Means by Reporting Category for Factor 6 | 125 |
| 7 | Means by Reporting Category for Factor 7 | 126 |
| 8 | Means by Reporting Category for Factor 8 | 127 |
| 9 | Means by Reporting Category for Factor 9 | 129 |
| 10 | Comparison for Familiarity with Published Guidelines (V53) | 146 |
| 11 | Comparison for the Fair Use of Examination Results (V60) | 147 |
| 12 | Comparison for the Formal Use of Results by School Administrators in Evaluation (V67) | 147 |
| 13 | Comparison for the Formal Use of Results by District Administrators in Evaluation (V69) | 148 |
| 14 | Comparison for School Administrators' Familiarity with Diploma Related Information (V54) | 149 |
| 15 | Comparison for District Administrators' Familiarity with Diploma Information (V55) | 150 |
| 16 | Comparison for the Informal Use of Results by School Administrators in Evaluation (V66) | 151 |

17 Comparison for the Informal Use of Results by District Administrators in Evaluation (V68).....151
18 Comparison for Teaching Assignments (V72).....152
19 Comparison for Transfer Requests (V74).....153
20 Comparison for Leadership Promotion (V75).....154
21 Comparison for Diploma Teaching Assignments (V78.).....155
22 Comparison for Teacher Stress (V76).....156
23 Comparison for Course Selection (V89.).....156

# CHAPTER 1

## OVERVIEW AND PURPOSE OF THE STUDY

The purpose of this thesis was to examine the perceived role of diploma examinations in Alberta high schools. The study was intended to gauge the attitudes and beliefs of school-based administrators and high school teaching staff on the use of diploma examination results. According to Loerke (1993), a number of researchers suggest that standardized test results are being used increasingly to measure teacher performance as well as student achievement. Diploma examinations in Alberta are considered standardized tests. They are designed to establish standards and measure the achievement of individuals and of instructional groups of students. A function of the diploma as a standardized test, is the ability to make valid comparisons based on provincial results. The establishment of provincial standards and the alignment of these standards with provincial goals requires input from a number of sources, including curriculum specialists, subject area teachers, field test information on student performance, and analytic resource staff. Educational standards are implemented through the curriculum; tests are used to ascertain if the standards are being met.

The heightened emphasis on accountability in education is evident in educational publications and the popular press. In the Calgary Herald, an article describes the role of testing. "Tests provide accountability, measure whether the curriculum is being taught, how well it's being learned and pinpoint students' shortcomings so that they can be properly addressed" (1998, A 22).

Ontario's version of a "literacy quiz" was piloted by grade 10 students across the province in October, 2000. The examinations were designed to

determine how well students were faring after a decade of education. In the *Ottawa Citizen,* education Minister Janet Ecker stated that,

> If there wasn't a problem, we wouldn't be doing the testing . . . in too many cases young people were coming out of high school without the proper literacy skills. Employers and universities and colleges were spending literally millions of dollars in remediation. (October 12, 2000, p. A5)

The test will be a diploma requirement for high school graduation in the 2001-2002 school year and is the latest attempt of the Ontario government to measure and improve the standards of the province's school system. The test is intended to be cross-curricular in nature; therefore, student preparation is not considered to be the sole responsibility of English teachers. Ministry initiatives to support improved literacy include: increased funding of early literacy programs; the introduction of specialized courses in the new high-school curriculum; and an investment in a teacher-adviser program to ensure students garner the support and advice they need in order to choose courses and plan their futures. Earl Manners, the president of the Ontario Secondary School Teachers Federation, cautions that, "what we're afraid of is that the test will be used to try to rank schools or teachers or students" (*Ottawa Citizen*, 2000, p. A5). Teacher testing is next on Ontario's political agenda.

Alberta Learning is also very interested in reassuring the public that it is serious about the quality of education, that it is constantly and consistently monitored, and that information is made available to the public. Legislation, such as the 1994 Freedom of Information and the Protection of Privacy (FOIPP) Act, is fundamental in terms of its mandate to make information available to the public.

Teacher evaluation is also a popular topic especially when linked to student achievement. According to Stake (1998) teaching is a very complex activity (Speech given November 19, University of Alberta). He believes that if

teaching is to be evaluated accurately, it requires that an equally complex evaluation mechanism must be in place. Teacher evaluation is also identified as one of the purposes of standardized testing in the Guidelines for Interpreting and Using the Results of the Diploma Examinations:

> Results from provincial assessments can assist teachers in their assessment of their own instructional practice and can assist others in the review of a teacher's instructional practice; however, results from provincial assessments shall not be used as the sole basis of evaluating teacher performance. (Alberta Education, 1998, p. 58)

Gary Mar, the former Alberta Minister of Education, was quoted in the Calgary Herald (1998) in an article as saying, "To stick your head in the sand and make excuses is not the purpose of the tests." The author described the "action not excuses" that the Minister expected from the Calgary Board of Education as it reviewed its provincial achievement test results and identified ways to improve them in the future. The author of the article implies that student results are directly related to teacher effectiveness. Teacher assessment and its connection to student results is one area of interest in this study.

The intent of student evaluation is to measure student achievement. Student achievement is not necessarily a measure of teacher effectiveness. But the two are inextricably linked and not mutually exclusive. According to Norman Henchey (1998) in an editorial in the Montreal Gazette:

> We do not compare hospitals in terms of the average blood pressure of their patients, however important blood pressure may be to [an individual's] health. A single indicator like test results in a few subjects offers only a sketchy context for analysis and interpretation and may be seriously misleading in judging the quality of a school...(1998, p.12).

The primary role of diploma examinations is to credential students: to determine which students meet the acceptable standard and to provide the necessary documentation of their achievement. Diploma results, with standards

that are statistically defined, are used as benchmarks against which student performance can be measured and compared.

As competition between individual schools and school districts increases, administrators are likely to rely increasingly on the use of standardized test results as a means to evaluate student learning and to monitor teacher effectiveness. Since the re-implementation in 1984 of the diploma examination program in Alberta, teachers and students have been under increasing scrutiny. The mandating of teacher growth plans and annual school reports, as well as the revamping of teacher assessment policies are but a few of the changes occurring in education in the province of Alberta.

## PURPOSE OF THE STUDY

The challenge of measuring school system performance has heightened emphasis on accountability by government, school boards, and other interest groups. Diploma examination results are being used to address this challenge. In 1993, shortly after diploma examinations were introduced, Loerke examined the uses of diploma examination results. He found that the practice of using achievement data to assess teachers did exist, albeit on a low level. He noted that the majority of teachers were unaware of the available information for accurately interpreting diploma results. Loerke suggested that this lack of information, in conjunction with the potential use of student results in formal and informal teacher assessment, contributed to the increased stress reported by diploma subject teachers. His study revealed that informed individuals were more likely not to use student examination results in teacher assessment.

The purpose of this study was to investigate the perceptions of Alberta teachers on the role of diploma examination results and the extent to which uses of diploma results have changed since Loerke's 1993 study. This research

aimed to replicate and extend that research project. This general purpose gave rise to five research questions:

1. How are diploma examination results used by Alberta teachers?
2. What is the relationship between the perceptions of various educational groups, such as school-based administrators versus teachers; diploma subject teachers versus non-diploma subject teachers; and among teachers in small, medium, and large-sized schools, regarding the role of diploma examinations?
3. How well informed are high school educators regarding the legitimate use of student test results?
4. To what extent are student test results being used as an endorsement of teacher competency?
5. To what extent have teachers' perceptions of the role of diploma examinations changed since the Loerke (1993) project?

Diploma examination results may be used to allocate teaching assignments, to set student participation rates for particular diploma courses, to set goals for professional growth plans, in teacher assessment, in school goal setting, and as an important part of school reports. One significant use of diploma results is in the assessment of curriculum delivery. Curriculum delivery refers to the dissemination of course specific knowledge and skills to students. Curriculum delivery encompasses many aspects of education including: student ability, teacher expertise, classroom instruction, learning facilities, use of technology, student evaluation, and teacher assessment.

## SIGNIFICANCE OF THE STUDY

In a premiere research study in Alberta, Loerke (1993) identified a number of literature references that supported his contention that it was inappropriate for administrators to use student test results to evaluate teacher

effectiveness. His review of the literature revealed that using student test results to assess teachers has questionable reliability, low correlation with other measures of teacher competence, and does not recognize that there are many variables related to student success. The use of results in formal teacher assessment allows individuals to challenge the assessment through district appeal procedures; however, the use of results in informal teacher assessment is a serious concern because these individuals have no such recourse.

One of the goals of the proposed study is to determine if these practices are more prevalent in Alberta today than in Loerke's 1993 project. This study also attempts to identify changes in perception and assessment practice regarding the role of diploma examinations. Significant professional, ethical, and moral implications regarding the use of student results are evident. It is important to remain vigilant of both positive and negative assessment practices in the field. This research study has the potential to contribute to the improvement of teacher assessment practices by assisting educators and policy makers in recognizing inappropriate uses for student examination results and by identifying appropriate strategies for linking assessment and accountability to improvements in student learning.

## SCOPE OF THE STUDY

This study replicates and extends Loerke's 1993 research. Gall, Borg, and Gall (1996) define replication as, "a study with a different group of research participants using the same or similar conditions, for the purpose of increasing confidence in the original study's findings" (p. 768). The research study is restricted to a survey of Alberta high school teachers and school-based administrators. Perceptions of the role of Grade 12 diploma examinations and

the relationship of students' results to teacher evaluation are the primary foci of this comparative research study.

The goals of Alberta Education's (1994) Meeting the Challenge: Three -Year Business Plan, the first in a series of three-year business plans, has resulted in the establishment of the first charter schools in Canada, economic down-sizing in education, subsequent drastic reduction in the number of school boards in the province, and a number of changes in the way public education is delivered. The changes in curriculum, resources, technology, economics, school organization, and student demographics since the 1993 Loerke study have likely had an influence on the perceptions of high school teaching staff regarding the role of diploma examinations.

Loerke used a questionnaire to investigate teacher perceptions of the use of diploma examinations in teacher evaluation. This study employs most of the original questionnaire items in order to enable comparison with the 1993 result, along with a number of new questionnaire items in order to elaborate on the various roles of diploma examinations.

This study addresses some of the missing aspects or shortfalls of the original study by utilizing teacher interviews as an alternate methodology. Interviews provided additional information, identified additional relevant factors, and offered insights which further support the quantitative research findings on the perceptions of administrators and teachers regarding the role of diploma examinations in Alberta high schools.

# CHAPTER 2
# LITERATURE REVIEW

This chapter presents a review of the literature related to the topic of standardized testing and the use of student examination results. The purpose of this chapter is to situate the research on the perceived role of diploma examinations within the context of the educational reforms in North America over the past twenty years. The review of the literature in this chapter is not exhaustive but selective, and highlights themes that provide a background perspective for the study and justify the selection of the research problems.

The review covers the following areas: (1) reform by comparison; (2) impact of measurement-driven instruction; (3) improving test scores; and (4) an accountability-driven reform. The chapter ends with a summary followed by a review of the 1993 diploma examination study.

## REFORM BY COMPARISON

The purpose of this first section of the literature review is to explore the education reforms deemed responsible for the current use of standardized, large-scale testing. Accordingly, attention is directed first to US reform strategies, and then to Canadian and Alberta initiatives.

The web of economic, social, and political forces in which the educational system is ensconced has resulted in major cycles of reform particularly during the past half century. In the 1950s, the race into space between the Soviets and the Americans provided an impetus for changing American schools and introducing methods, such as inquiry teaching, in education. In the 1960s and 1970s, North America struggled to address economic and social costs of poverty and racism, resulting in access and equity in education policies with a strong

focus on curriculum content. The 1970 reforms spotlighted assessment of student performance. The period from 1980 to the present delineates a generation of educational policies focused on enhancing student performance, commonly referred to as the standards movement (Murphy, 1998; Stake, 1998). In the 1980s and 1990s, policy makers responded to the rise of the global markets with the advent of the excellence era in education. The 1980s are characterized by outcome-based reforms, and the 1990s with results-based models. In results-based reform, student performance goals are made more explicit so that testing can be more precisely focused, along with efforts to align curriculum with testing (Stake, 1998).

Underlying virtually all secondary education innovations of the 1980s was the conviction that standards had slipped, expectations had eroded, and the system as a whole had grown sluggish and inefficient (Finn, 1991). The height of US public and political attention on education occurred in 1983 with the release of a report by the National Commission on Excellence in Education entitled *A Nation at Risk: The Imperative for Educational Reform* (Berliner and Biddle, 1995). The report's message was that schools and universities were responsible for the decline in America's greatness in the world (McConaghey, 1983). Education reforms that started in earnest in the 1980s and continue into the present can be traced almost directly to this report. War was declared on mediocrity (Berliner and Biddle, 1995). The result was the heightened expectations in high school graduation requirements and the adoption of more rigorous and measurable standards for academic performance. This US reform agenda has driven many industrialized nations to implement educational reforms because of a perceived crisis in accountability (Mawhinney, 1995; McEwan, 1995; Cibulka and Derlin, 1995; Earl, 1996; Macpherson, 1995).

A 1989 *Statistics Canada* survey of Canadian literacy skills formed the basis of the position held by the Economic Council of Canada that Canadians were not meeting everyday learning demands. The Council predicted that a third of the student population was at risk if the school system did not improve. In the report, *A Lot to Learn*, the Economic Council of Canada gave schools a failing grade, largely based on student achievement on international mathematics and science tests. Educators cautioned the Council that the tests were not curricular based and that the populations taking the test were significantly different making student achievement outcomes problematic. Schools were blamed for graduating non-literates. Barlow and Robertson (1994) refuted this "myth". They contend that the *Statistics Canada* study was based on an older population, many of whom were educated in another time or place, were not born in Canada, and did not speak English as a mother tongue.

Along with concerns about rising illiteracy, concern with increasing drop-out rates was expressed in virtually every public statement on Canadian education in the late 1980s and early 1990s. Student drop-out rates were commonly said to be 32 percent. The Conference Board of Canada predicted that these high rates would have huge consequences for Canadian society. In 1993, *Statistic Canada* revised reduced its figure on student completion to 18 percent. Apparently, the previous rates did not take into account students who did not graduate in the expected year, or students who graduated from a different school than the one in which they started. The 1993 *Statistic Canada* report showed that Alberta had the highest retention rate at 86% (Barlow and Robertson, 1994).

The ATA's 1992 task force, established to study teachers' experiences with education reform in Alberta, published a number of documents in a series called, *Trying to Teach, Western Report*. Although teachers appeared to accept

the principle of accountability, they viewed the imposition of standardized, external testing as a diversion of time, energy, and education funding. Teachers voiced a suspicion that accountability systems, such as the diploma examinations program, were not driven by an educational reform agenda, but rather by a political agenda designed to alleviate criticism of education and placate public concerns. In Canada, as in the US, the public appeared to have, "settled on better student performance as the desired means to attain a strong economy, vibrant democracy, and satisfied citizenry" (Murphy, 1998, p.441).

The excellence reform movement is now starting its third decade on both sides of the Atlantic. Categories of change that have become the hall- marks of the excellence movement include various restructuring initiatives, strategies for extending school choice, and systems of standards and accountability that bear consequences (Mawhinney, 1998). It is important to contextualize education reform movements. The connection between education policy and the prevailing world view influences how policymakers interpret problems and the ways they seek possible solutions.

The focus of this study is on the establishment of standards and resulting accountability that bear consequences, namely diploma examinations. Historically, the concept of accountability has reflected a relationship between a steward or provider of goods and services and a patron or agent with power to reward, punish, or replace the provider (Kirst, 1990). For schools, accountability can be defined as the process by which school districts and provinces, or other constituents such as parents, attempt to ensure that schools and school systems meet their goals (Rothman, 1995). For many people, school reform means taking the responsibility for goal setting away from teachers and externally assessing common student accomplishments (Stake, 1993). Two very different theories of school reform are working in parallel and sometimes at cross

purposes. One theory focuses on tightening the controls: more standards enforced by more rewards and more sanctions. These reformers would improve education by developing more tests and tying funds to schools' test scores. A second theory attends more to the qualifications and capabilities of teachers and to developing schools through changes in teacher education, licensing, and certification processes. Among the major strategies are professional development, efforts to decentralize school decision making, changes in local assessment practices, and the development of networks among teachers and schools (Darling-Hammond, 1992).

Assessment-driven reform is not new and can be traced back to the origins of public education in the US. As early as 1845, Horace Mann recognized that, "school-by-school test results would provide political leverage over recalcitrant head masters" (Ramirez, 1999, p. 205). Assessment is a synonym for evaluation as it requires both a description and judgment regarding whatever phenomenon is being assessed. Assessment plays a pivotal role in standards-based reform. The logic of standards-based reform holds great appeal with policymakers: standards provide a basis for instruction and student learning; the assessment of the standards provide targets for district schools, teachers, and students; assessment results help identify the system level that needs to improve; results are used by professionals and students to meet goals; and sanctions and rewards are used to encourage improvement (Herman, 1997).

The underlying assumptions that drive the current standards-and-assessment-driven reform movement are familiar and disturbing. What is upsetting are the following assumptions: that students are unmotivated and require immediate consequences to their learning; that teachers are inadequately skilled or lack the motivation to inspire students to higher levels of

learning; that local school boards and superintendents don't know what their students should be learning or to what degree they should be learning it; and that accountability through testing will pressure the system to improve (Ramirez, 1999).

Shepard (1994), Meisels (1989), and Bredekamp and Rosegrant (1992), suggested that assessment can serve several beneficial purposes such as informing instruction and guiding decisions about classroom teaching, identifying children in need of intervention or special services, and evaluating how well a program is meeting its goals (as cited in Donegan, 1998). The professional literature also suggested that the conclusions drawn and the decisions made from the test results should be limited to the intended purposes of the assessment measure.

It should be noted that much of the educational research is based on standardized large-scale examinations which have different purposes and varying stakes than the Alberta diploma examinations. Included in the literature are many different types of examinations, such as minimum competency and achievement examinations, with varying stakes from low to high, depending on their significance in student and staff evaluation. Findings from the United States and other countries should be viewed in context. Many differences in examinations are related to factors such as cultural disparity, educational policy, and funding equity.

## Introduction of Large-Scale Testing

Testing is a single way of getting a representative sample, thus a limited amount of information about what a student knows or is able to do in reference to a specified domain of knowledge or behavior. The multiple-choice format test was first introduced in 1915 by Frederick Kelly, then mobilized by Arthur Otis, a

psychologist at Stanford University, who is deemed the father of multiple-choice testing. Because of its low cost and efficiency, multiple-choice tests quickly became a predominant method of classroom assessment. High stakes were attached to early large-scale assessment. The term "high stakes" refers to tests with severe consequences such as the use of test results to rank students, schools, and districts in the media (Popham, 1987). One such test, the Scholastic Aptitude Test (SAT), was first administered in 1926 to select college entrants (US. Office of Technology Assessment, 1992). Today, the SAT scores are frequently cited to make judgments about the state of education in America– well beyond the test scores' original intent. Concerns regarding the appropriate use of assessment data, the impact of assessment on instruction, and fairness issues, have intensified since the implementation of the SAT.

The emergence of systematic external-testing changed the nature of assessment from a tool for teachers to an instrument of public policy to bring about change (Asp, 1998). How to balance the competing purposes of assessment to improve instruction and to increase student achievement has been a major issue since the inception of large-scale assessment. Tension can be found between using a single assessment system both to promote instructional improvement and to promote accountability (Fuhrman and O'Day, 1996). An assessment that attempts to perform too many functions–student diagnosis, curriculum planning, program evaluation, instructional improvement, accountability, certification, and public communication– will inevitably do nothing well (Linn and Herman, 1997). Massive increases in large-scale testing in the 1970s and 1980s have identified a related area of reform–the need to address teachers' lack of knowledge about high quality assessment practices (Stiggins, 1991). Teacher preparation programs provided little or no training in formal assessment.

For the past 30 years, assessment has been a significant means of quality control and an instrument of educational reform (Stake, 1998). It has become the fundamental tool for educational reform in an era where accountability is not an option. Both the public and policy makers view assessment as the engine of reform which can improve education rapidly and for relatively low cost (Popham, 1998; Linn, 1998). The criteria for evaluating the quality of assessment and assessment systems should include the influence of assessment on learning, teaching, and the school as an organization (Sykes, 1997). No statistical or psychometric magic can produce these criteria because they are ultimately human constructions. It is incumbent on test developers to inform policymakers, practitioners, and the public about appropriate uses and limitations of an assessment tool. School districts need to clarify what targets they truly value in order to help teachers see the link between classroom instruction and student scores on large-scale assessment (Asp, 1998).

As an education reform strategy, testing comes under the broad heading of accountability–ways of furnishing parents, policy makers, and educators with accurate information about the efficacy of their efforts. It is a powerful tool since what gets measured gets attention (Finn, 1991). Notwithstanding palpable consequences may be linked to results in high-stakes testing, it remains a relatively inexpensive and easy reform to mandate. Many politicians and policy makers today link school accountability and school performance. As in the corporate world, they assume that strong external accountability will impel schools to improve student achievement.

## The Politics of Accountability

Since the early 1980s, a growing trend toward reform by comparison is evident in Alberta. Advocates of large-scale testing believe that,

> Measurement is one of the best ways for a school to see where it is in fact strong or in need of improvement. Visible, tangible indicators can reveal not only who is and isn't getting a good education; they can also reveal areas of program strengths and weakness, which in turn create opportunities for improvement. (Schmoker, 1997, p. 132)

Reinventing collegiality is another challenge of our time. Teachers, in many cases for the first time, are asked to be thinking contributors who can generate solutions to emergent problems and obstacles (Schmoker, 1997). The complex problems of translating curriculum into practice is greatly complicated by the overlapping expectations and political influences of administration, school boards, test developers, professors, the public, and students (Hart, 1988).

"Declining education is ultimately a responsibility not simply of each individual school, but of the broader policy that funds and grants schools the authority to operate" (Newmann, Rigdon, and King, 1997, p.42). Agencies such as Alberta Learning and school boards have legal and political responsibility for the quality of education in Alberta. The School Act provides the legislative framework for sustaining and developing Alberta's education system. The five underlying principles of the Act include access to quality education, equity, flexibility, responsiveness, and accountability. Under the heading of accountability, the School Act states that all those involved in making decisions about educational matters must be accountable for their decisions, choices, and results.

The public demand for stringent quality reviews to evaluate the multi-billion dollar education investment has the greatest possible impact on student learning (Raham, 1999). Using student scores on provincial tests to make judgments about the quality of education provided in a particular school or district has evoked changes in what happens in the classroom (Murnane and Levy, 1996). Assessment and accountability are inextricably linked. Assessment is the gathering, organizing, and reporting of information.

Accountability uses this data to make judgments about where to adjust behaviors in a constant cycle of assessment, analysis, and action to improve future performance. Because learning is complex, multi-dimensional, and dynamic, achievement is but one indicator of successful learning and growth (McEwen, 1998). Test scores alone do not provide sufficient information for school success. A broad range of indicators is needed, including parent and student satisfaction levels, attendance rates, school leadership, community involvement, and valuable student learning which cannot be captured on examinations. Complementary use of internal and external evaluation can point schools to the path for necessary change to improve student outcomes (Marchesi, 1998).

Impact of Testing on Teaching Practices

As in Alberta, British Columbia had province-wide, school-leaving examinations that had been an integral part of education until 1973 when they were abandoned for a period of ten years. In British Columbia in 1983, and in Alberta in 1984, the governments reinstated a system of centrally set and marked provincial examinations. The Minister of Education in Alberta, David King, announced the reinstatement of the diploma examinations in response to public concerns about the declining quality of education (Samiroden, 1991).

The ten year period with no final province-wide examinations followed by their reintroduction provided an opportunity to study the impact of the effects of large-scale testing in British Columbia (Wideen, O'Shea, and Ivany, 1992). The BC study revealed that grade 12 final examinations had become more than mere indicators of student performance and that their effect was felt beyond the classroom, raising issues of a broader educational nature. The examinations' main role was one of comparative reporting of school and district results.

Results were used to discourage marginal students from taking certain grade 12 classes in order to maintain high averages for schools and districts. The practice of replicating the final examinations in grades 10 and 11 to parallel those in grade 12 was common, with the rationale that such examinations were implemented to "ready" these students ready for the grade 12 experience. There was evidence of subtle district and school decisions involving the selection of teachers who could produce the best results, although neither districts nor schools produced any evidence of written or spoken policy regarding the implications of large-scale testing (Wideen, O'Shea, and Ivany, 1992).

The Wideen, O'Shea, and Ivany (1992) study revealed that grade 12 teachers are much more influenced than other grade level teachers by external factors of examinations and curriculum guidelines and saw large-scale testing as having a much greater impact on their teaching. A typical grade 12 teacher was influenced almost entirely by the curriculum and examinations. Diploma examinations were used as a means to motivate students, review material, ensure student attentiveness, diagnose teaching, and ensure external accountability. Time constraints reduced the number of laboratory projects done in science courses and increased the tendency by diploma course teachers to lecture and to spend valuable classroom time reviewing before examinations. The reduction in the number of topics covered, especially if they were not tested on examinations, caused a narrowing of the curriculum. Teachers expressed concern about these shifts in classroom practices that resulted in the concentration on teaching facts, content oriented instruction, and a reduction in the development of a sense of curiosity and a sense of social responsibility in students.

The loss of teacher autonomy and creativity, along with the increased psychological pressure associated with final examinations that are used for comparison purposes, added to the list of concerns. Teachers sensed that their teaching was assessed largely on the basis of student performance which they felt was not always an accurate reflection of how well they had taught a particular course. Concerns were also expressed about the quality of examinations as fair determinants of students' knowledge and a need for a better understanding of the unintended effects of external testing on students. From the analysis, it was evident that the impact of the final examinations went well beyond what may have been originally intended by the government who initiated them (Wideen, O'Shea, Ivany, 1992).

Diploma Re-implementation in Alberta

In an effort to improve the perceived erosion of standards since the withdrawal of diploma examinations in 1973 and to address public accountability issues, Alberta Education embarked on a series of reforms with links to student assessment. Policy makers implemented the Achievement Testing Program for grades 3, 6, and 9 in core subjects in 1982, followed in 1984 by the re-introduction of the grade 12 diploma examinations in seven academic subjects. Not only were external examinations reinstated, the number of specified credits for an Advanced Diploma and General Diploma were also affected. The specified advanced and general diploma requirements in 1984 went from 58 and 45, respectively, to 76 and 62 by 1992 (McEwan, 1998). Along with these assessment and accountability reforms, many of the diploma courses underwent significant curriculum changes to enhance the content and increase rigor.

The Alberta Teachers' Association (ATA) and others expressed concerns that a single diploma examination cannot be used for the numerous roles the

government outlined in 1983. The goals of the diploma examinations were to: clarify the goals of education, perform a bench marking function, evaluate programs, establish standards, maintain standards, certify levels of achievement, certify minimum competency, predict student success in post-secondary institutions, enhance public confidence, and serve as the major criterion for a high school diploma (ATA News, 1983a, 1983; Krembil, 1987; Linn and Herman, 1997). Concerns were also raised about the possible reduction in the pursuit of courses in fine and practical arts by high school students, given the number of specified core courses required for graduation (ATA News, 1983a). Despite these issues, the ATA supported the re-implementation of diploma examinations at the time due to concerns about an apparent lack of standards and the erosion of public confidence in public education. The ATA was pleased with the restoration of examinations because they provided an external frame of reference for student evaluation to address the earlier concerns (ATA News, 1983c). The two conclusions drawn by the ATA's *Task Force on High School Student Evaluation* were that the public wanted some kind of provincially administered final examinations at the high school level and that students should receive a high school diploma only if they pass such examinations. The task force also found support for the belief that teachers can best evaluate student progress (ATA News, 1983b). A number of criticisms of the diploma examination program propelled the ATA to commission a study to explore the negative allegations and to assess the impact of diploma examinations designed for school graduation. Calder (1990) concluded that,

> Although there is no consensus, and teachers see many ways in which the examinations negatively affect the teaching-learning process, they generally favor the retention of diploma examinations. The examinations are seen more as fulfilling a political evaluative function than being of educational value. Although teachers report more negative impact of the examinations, they seem to be reluctant to dispense with them. (p.3)

The ATA has since provided support for the diploma program by way of teacher participation in every aspect of examination development and approval. There are however, many concerns regarding the continued use of these examinations. The agency that administers the test assumes a great deal of power over the schooling processes. The test results can impact decisions that affect individual life changes such as high school graduation requirements, teacher tenure decisions, and the allocation of education funding (Madeus, 1981). One of the greatest concerns of teachers regarding the diploma examinations is the possible misuse of test results in the evaluation of teachers (Calder, 1990b).

## IMPACT OF MEASUREMENT-DRIVEN INSTRUCTION

The purpose of this section of the literature review is to investigate the impact of measurement-driven reform on instruction and on the behavior of students, teachers, and administrators. Setting assessment targets, the use of test scores, ranking schools, media impact, and test quality, give rise to issues that collectively challenge the value of large-scale testing programs.

### Instruction Diverted

When educational policy decisions are made on the basis of external tests results, administrators, teachers, and students take the examinations seriously and modify their behavior and attitudes accordingly (Madeus, 1981). Alfie Kohn, in his latest book, The Case Against Standardized Testing: Raising the Scores, Ruining the School (2000), states that standardized tests tend to ignore the most important qualities of a good learner such as initiative, conceptual thinking, commitment, judgment, and ethical reflection. Teaching to the test, narrowing of the curriculum, tapping only lower-order thinking skills,

increasing administrative costs, disadvantaging low achievers, and test bias that is unfair to minorities, are among the key objections of the education community to standard assessment programs (Calder, 1990; Raham, 1999). While many of these concerns can be overcome or addressed through more sophisticated assessment practices, they remain common obstacles to system measurement (Phelps, 1999). The concerns also require test developers to be vigilant about instructional implications, to promote test validity, and to build safeguards to minimize potential risks (Krembil, 1987).

The message from the opponents of large scale assessment is loud and clear. Standardized tests divert what is taught. Examinations test only high priority content (Stake, 1998). They have the deleterious effect of "narrowing the curriculum" which promotes surface learning and memorization and focuses on the parts of the curriculum most likely to be tested. Indeed, the tests become the curriculum (Samiroden, 1987; Bracey, 1987; Phelps, 1999; O'Shea and Wideen, 1993; Meaghan and Casas, 1995; Ramirez, 1999). The pressure to improve student scores with each administration not only becomes oppressive but more importantly diverts teachers from considering the other, possibly more educationally valuable, objectives of the program (Samiroden, 1987; McConaghy, 2000). Emphasizing the measurable will force a shift towards areas that are most easily measured with a reliance on standardized tests to the exclusion of all other measures (Kohn, 1992).

Teachers in Samiroden's (1988) study of the effects of diploma examinations on science instruction in Alberta, reported that the examinations curtailed learning activities such as laboratory exercises, projects, library research assignments, group work, brainstorming sessions, field trips, enrichment activities, science fair participation, and debates in class forums, in order to allow more preparation time for material that would be on the

examination. The study also found that Alberta science teachers were strongly influenced in their approaches to teaching and classroom testing by the content and structure of the externally developed diploma examinations. Much like the findings of the British Columbia study (Wideen, O'Shea, Ivany, 1992), teachers have spent more class time lecturing, testing, and reviewing since the 1984 re-implementation of the diploma examinations. Diplomas have become a way for administrators to appraise the performance of their teachers in relation to other teachers in their jurisdiction and in the province-at-large (Samiroden, 1987). Diploma test results have been used to compare teachers, schools, and school jurisdictions. Such comparisons, and indirect evaluations, have induced teachers to teach toward and to emphasize the content that the diploma examinations test (Samiroden, 1988).

Others believe that if the local curriculum overlaps with the objectives of a standardized test, then teaching to the test is inevitable and desirable (English, 1992). Diploma examinations promote curricular alignment, which covers a wide range of understandings and activities that involve the relationship between instruction and achievement, legitimate issues of goal setting, textbook selection, and curriculum design (Pipho, 1988). Teachers are positively affected by external examinations in several ways. Advocates for measurement-driven instruction argue that if tests measure important skills and have sufficiently high stakes, they will serve as "instructional magnets" thus dramatically improving the efficiency and effectiveness of instruction (Popham, 1987). Teachers may follow the curriculum more closely, develop teaching methods which enhance student learning, and use external examinations to improve their own methods of evaluation (Swanson,1994; O'Shea and Wideen, 1993; Bishop 1998b). The benefit of large-scale assessment is that it focuses attention on the areas that are valued and important and as such can provide evidence of improvement. A

successful assessment program measures those important outcomes, provides a common metric for all students, helps create a common expectation for student learning, and involves teachers in all aspects of the examination process from its development and validation to the marking of student work. Teacher involvement provides invaluable professional development in devising pedagogically sound curriculum and effective assessment strategies. A successful assessment program also provides information for a variety of audiences on an ongoing and timely basis (McEwen, 1998).

The list of criticisms is long and varied. Robert Stake (1998), an opponent of large-scale assessment, believes that:

> the validity of measurement of achievement is not the same as validity of those same test scores as an indicator of quality of teaching and learning conditions. Teaching can be changed in a number of ways within a school or classroom without change in achievement terms. Using those scores as a measure of school improvement has not been validated. No accumulation of evidence shows assessment to be an indicator of good schooling. ( p.5)

Examinations promote a focus on goals rather than on process (Swanson, 1994) and promote various methods of test score inflation (Phelps, 1999). Teachers and administrators may devise ingenious ways to raise test scores without necessarily upgrading skill levels (Meaghan and Casas, 1995). There is a lure towards data pollution and falsification of scores (Stake, 1998). The demoralization of teachers and negative views of self and school are also among the criticisms (Samiroden, 1991; Stake, 1998). A tendency to blame poor results on poor instruction which is redirected toward students when it should rest with the profession and the authorities (Stake, 1998). Some believe that large-scale testing is too costly (Phelps, 1999a) and it withholds needed funding from student education (Stake, 1998). Grades spoil students' relationships with each other. When students are not just rated but ranked, the message is that the purpose of education is not to learn, or even to perform well,

but to defeat others (Johnson and Johnson, 1989; Kohn, 1992). Some suggest that schools often exist for the purposes of comparative credentialing rather than for the legitimate goal of helping all children learn (Kohn, 1998).

Students' grades from externally developed educational tests have been used for a variety of assessments other than that of student achievement. Results are used for a variety of purposes, including the evaluation of the effectiveness of teachers, curriculum, educational systems and programs, the identification of trends in educational achievement, the comparison of the standings or effectiveness of schools, school districts, provinces, or the nation with respect to educational excellence, and to assist in curriculum planning and policy making (Broadfoot, 1984; Linn, 1986; Nickerson, 1989; Snow, 1989; Samiroden, 1991; Jones et al., 1999). Conclusions drawn from standardized test results extend well beyond the intended purpose of the instruments for valid group comparisons. Pressures on local schools and classroom teachers to improve test scores have resulted in pushing down curriculum expectations to earlier grades, reducing curriculum to those skills amenable to testing, and spending more instructional time on the teaching of test-taking skills (Nolen, Haladyna, and Haas, 1992; Urdan and Paris, 1994, McAfee and Leong, 1997, Meisels, 1989, Nolen et al., 1992, Seefeldt, 1990, in Donegan, 1998). Many, however, believe that the benefits outweigh the costs (O'Shea and Wideen, 1993). Benefits include diagnostic and evaluation information, positive classroom outcomes such as curriculum alignment with standards, and positive products of the assessment processes such as clear standards, better public understanding, and teacher edification (US General Accounting Office, 1993).

Madeus (1989) summarized the consequences of measurement-driven instruction in a high-stakes testing situation in terms of the following general principles.

1) If students, teachers, or administrators believe that the results are important, whether true or false, they are extremely powerful in defining the reality of how a test is used.

2) If important decisions are presumed to be related to test results, then teachers will teach to the test, thus effectively narrowing the curriculum.

3) In every setting where a high-stakes test operates, a tradition of past tests develops, which eventually de facto defines the curriculum. Proponents of measurement-driven instruction agree strongly that if the skills are well chosen, and if the tests truly measure them, then coaching is perfectly acceptable (Millman, 1981 and Popham, et al 1985, in Madeus, 1989).

4) Teachers pay particular attention to the form and format of the questions on a high stakes test and adjust their instruction accordingly.

5) When test results are the sole or even partial arbiter of future educational or life-choices, society tends to treat test results as the major goal of schooling rather than as a useful but fallible indicator of achievement.

Effect on the Behavior of Students, Teachers, and Administrators

In a recent Canadian study, Bishop (1998a) concluded that schools in diploma examination provinces scheduled more hours of mathematics and science instruction, assigned more homework, had better science laboratories, were significantly more likely to use specialist teachers, and were more likely to hire teachers who had studied their teaching subject in college or university. The diploma examinations did not affect the hours in the school year, class size, and teacher preparation time, the number of computers in the school, and the reporting of discipline problems. In provinces that had political support for

diploma examinations, teachers administered more tests and quizzes more frequently and principals were significantly more likely to report student absenteeism problems (Bishop, 1998b). Despite an enormous body of empirical research, there is currently little consensus about whether additional education spending will improve student test scores, the most commonly used measure of student learning (Ludwig and Bassi, 1999).

Evidence from both Canada and abroad suggests that curriculum-based, externally-set exit examinations increase student achievement. International testing evidence suggests that students from countries with medium and high-stakes exit examinations outperform students from other countries at a comparable level of economic development (McEwen, 1998b). Diploma examinations appear to improve both the quality of instruction and the quantity of student achievement (Bishop, 1998a). The diploma examinations are but one factor contributing to accountability in education. The challenge for this millennium is to design and implement truly effective accountability systems that change the internal dynamics of schools, address the will and capacity of teachers to use performance data to improve their practice, and engage the whole community in supporting improved student learning (Raham, 1999).

Setting Targets

Governments can influence education system performance by analyzing strengths and weaknesses and identifying improvement targets. Alberta Learning's well established *Three-Year Business Plans* (2000/2001) outline the comparative progress of the system over the previous years on a wide variety of indicators. The report announces specific provincial targets and key measures for three consecutive years. The primary goal of the current three-year plan is to focus education on what students need to learn for further studies, preparedness

for work, and citizenship. The goal is also for all Alberta students to achieve acceptable provincial and national learning standards and to perform well in international comparisons. One of the key measures of this goal are the results on diploma examinations. Schools and boards are required to develop annual education plans and teacher growth plans that are aligned with provincial targets and to report on their progress (ATA News, 1999a; Teacher Growth, Supervision and Evaluation Policy, 1998). This act of quantifying goals has focused attention, resources, and efforts on the results. Alberta's consistent rating at or near the top in student achievement on national and international assessments, respectively the School Achievement Indicators Program (SAIP) and the Third International Mathematics and Science Study (TIMMS), may be evidence of the success of this goal setting process (McEwen, 1998b).

Although standard setting is a critical component in many testing programs, evaluating how well a standard-setting procedure is working is difficult. Judgment is a large component in all standard setting, particularly in defining the performance standard (Popham and Scriven, 1978, in Kane, 1998). A certain amount of variation in student means and percentage of students achieving standards on diploma examinations from administration to administration and from year to year, is common. In supporting the goals of communicating high standards, and improving instruction and assessment in schools, the test items used in major administrations of the diploma examination (January and June) are made public and are not reused. The challenge then is to produce equivalent exams with different items each administration. Currently, Alberta Learning has only limited evidence as to whether any variations are a result of changes in student achievement, the variation in the ability of students taking the course, measurement error, or a variation in the standards of the examination (Wasserman, 1999).

Uses of Test Scores

A school district that bases its accountability system on test scores alone is analogous to a physician who evaluates physical health based upon a single indicator such a body temperature, while ignoring the other medical indicators that any reasonable physician would regard as essential to a competent diagnosis (Reeves, 2000). Diploma examination scores should not be used to indicate how good or bad a school is; rather, they should be used to make the comparative interpretations that they were intended to provide. Parents and educators often ascribe far too much precision and accuracy to students' scores on standardized tests. What educators need to do is spend some quality time with standardized tests (Popham, 1999). Teachers need to scrutinize the tests items one-by-one to ascertain what they are really measuring. In order to improve student learning, Alberta Learning, school authorities, and schools analyze, interpret, use, and communicate the results of provincial assessment. The results are meant to help identify areas of strengths, areas in need of improvement, and progress made towards the achievement of improvement goals or targets. The careful interpretation of results from diploma examinations informs decisions about how to improve student learning as outlined in *Accountability in Education: Use and Reporting of Results on Provincial Assessments, Policy 2.1.3* (1998b). Familiarity with the guidelines for interpreting diploma examination results is crucial to avoid potential misuse of student results. The School Act outlines the roles and responsibilities of teachers, principals, the school council and the treatment of student records.

Educators should definitely be held accountable. If educators accept the position that standardized test scores should not be used to measure the quality of schooling, then they must provide other, credible evidence that can be used to ascertain the quality of schooling (Popham, 1999). Teachers need a working

knowledge of: the relationship between diploma examinations and the Program of Studies; diploma examinations and ongoing classroom assessment; population size and the meaning of results; examination marks and school-awarded marks; results from a single administration and results over time; participation rates and results; achievement of standards and achievement of averages or means; and examination results and target setting. To this end, *Interpretation of Results Workshops* for diploma examination subjects are hosted by Alberta Learning for various stakeholders throughout the province to facilitate the open-communication policy goal.

Student Results and Ranking

The Freedom of Information and Protection of Privacy (FOIPP) legislation came into effect for Alberta school boards on September 1, 1998. The FOIPP Act aims to strike a balance between the public's right to know and the individual's right to privacy, as those rights relate to information held by public bodies in Alberta. Ultimately, the Act allows parents and students broader rights of access to information and obligates schools to protect individual privacy. Parents are becoming better informed about how well schools are performing and are demanding more information. In fall 1998, Alberta Learning posted public versions of the school and jurisdiction multi-year survey reports on its Website. Direct access to these and other reports allow parents to base their decisions about their children's education, in part, on how well students do on provincial assessments (McEwen, 1998). FOIPP has expanded the media's access to education related information and has contributed to raising the stakes attached to assessment programs. The posting of school-by-school test scores on the Alberta Learning Website is a result of FOIPP and the Alberta

government's open-communication policy which provides access of public information to private groups.

The use of student examination results to rank schools is not unique to Alberta. The League Tables in Great Britain and the annual publishing of SAT scores in the US have a long history as ranking instruments. The publishing of school rankings has the potential to increase pressure on teachers, students, and administration. The *Calgary Herald* and the Fraser Institute publish annual rankings of schools in terms of selected diploma test results. School Works! Inc. has re-released its provocatively titled report, *Top 200 Failing Alberta Schools*, determined by ranking schools according to their scores on provincial achievement and diploma examinations. School Works! Inc. rankings are seen by many as evidence that children in the province are being shortchanged educationally, while others see the rankings as the most extreme example of a growing pressure to improve test results. Critics argue that the rankings are based on incomplete data and use test scores for purposes for which they were not intended. Moreover, the ranking of schools exaggerates small differences, creates distinctions where they don't really exist, and have the potential to increase misunderstanding and anxiety (Hoffman, 1999). As well, the school ranking changes, in some cases drastically, as the criterion changes (Guskey and Kifer, 1990). When a school has a small number of students writing these examinations, one or two struggling students can make the difference between a good ranking and a bad one (Hoffman, 1999).

In the *Edmonton Journal* on July 16, 2000, ATA president Larry Booi observed that the obsession with test scores bears little relationship to real education. "It's almost as if what we were saying is we don't need to feed the cattle more–we just need to weigh them more often," states Booi (also in Jones et al., 1999). Learning Minister, Dr. Lyle Oberg suggested that ranking schools

paints a false picture and to simply say that one school is better than another school because of a test score is wrong (July 16, 2000).

In order to rank schools or school districts, one must assume that both the criteria used and the technique employed are defensible (Guskey and Kifer, 1990). To base ranking or school effectiveness on a single measure – such as results on diploma examinations – without regard for student factors is to generate potentially misleading information (Kendall, 1995; Fetler, 1991). A vast body of research (Chwialkowska, 1999; Jencks, 1972; Wang, Haertel, and Walberg, 1993; Lissitz, 1997; Gary, Jesson, and Goldstein, 1995) reveals that children from socially disadvantaged areas tend to have lower examination scores and test results than those from socially advantaged areas. Comparisons are inherently unfair, regardless of the nature of the testing. The background that students bring to school may raise or lower the quality of education by the mere concentration of the student peer group. Preparation, standards of performance, or attitudes for learning, in general, are associated with family socioeconomic status. Family composition is by far the strongest predictor of school performance (Bankston and Caldas, 1998; Caldas and Bankston, 1997; Rothstein, 1998). One example of family impact can be found in the Statistics Canada study *Determinants of Post Secondary Participation*. This study reports that parents with a university education were three-and-a half times more likely to have their children attend university than children whose parents had only completed high school. The children from both groups had the same grades in high school, worked the same number of hours at part-time jobs, and had the same level of high school science, mathematics, and reading skills (Chwialkowska, 1999). Decades of research indicate that socio-economic levels (Jencks, 1972) and differences in basic ability (Wang, Haertel, and

Walberg, 1993) are the primary covariants, if not determiners, of school success (Lissitz, 1997; Gary, Jesson, and Goldstein, 1995).

Student achievement is the product of many socializing institutions, not school alone. Increased sophistication about these non-school influences does not imply iron determinism, but it does mean that unadjusted scores shed little light on the quality of education. Interpretations of assessment results should take account of the backgrounds and learning experiences of the students (Principles for Fair Assessment in Canada, 1993).

A comparison of school averages, or distributions, tells us nothing about the relative achievements of different types of students within the schools. Group comparisons tell us nothing about the progress and educational needs of individual students (Hoffman, 1999; Reeves, 2000). Two schools may achieve the same average results by quite different means. Consequently, schools which perform well relative to other schools for the average students in the population, may perform less well for disadvantaged or advantaged students (Cuttance, 1988; Randenbush and Bryk, 1987). Provincial tests provide teachers with valuable feedback about how their own standards compare with those of their colleagues. Provincial assessments can provide an overview of trends in a certain subject area in a district or school, but in terms of any given child's achievement they are of limited value (Hoffman, 1999).

Unlike automobile plants, schools cannot control the variables that affect test data and outcomes (Schmoker, 1996). Prior knowledge is one such variable. Jonassen and Gabrowski (1993) define prior knowledge as, " the knowledge, skills, or ability that students bring to the learning process" (p. 417). It is also important to recognize that prior knowledge can refer to correct understandings or incorrect misunderstandings, often referred to as misconceptions. Research indicates that it is difficult to overestimate the

contribution of an individuals' prior knowledge. Prior knowledge is an essential variable in learning (Dochy, Segers, and Buehl, 1999).

Media Impact

The most pervasive source of high stakes pressure identified regarding test results is media coverage (Shepard, 1990). The media contribute to the problems of the increased focus on standardized achievement performance by publishing ranked lists of schools and districts without including important external influences such as demographic characteristics (Shepard, 1989; Nolan, Haladyna, and Haas, 1992). If it "bleeds, it leads" is too frequently the criteria for printing stories that news reporters think the public will find interesting. Articles that are critical of public schools or have some "scent of blood" often get printed. According to Shepard (1990), on educational issues, the press is biased and covers the negative side of the news stories more diligently than the positive side. They present a too simplistic and incomplete view of educational problems and issues, and display a lack of understanding of the complexity of school life. Berliner and Biddle (1999) add that the media demonstrates an appalling lack of understanding of statistics and social science research without which reporters cannot properly interpret the large amount of data the educational system produces. When it comes to sports, we expect the best commentators to look beyond the data and provide insights based on observations, description, and qualitative understandings that extend beyond numerical explanations. Commentators are expected to demonstrate a thorough understanding of the nuances of the game (Reeves, 2000). The analysis of educational accountability should be taken as seriously by the media and the public as analysis of last weekend's sports games.

The media were cited as the most useful source of information by respondents to an Alberta Learning survey (McEwen, 1998). Alberta newspapers publish more than 1200 articles on education per year. Alberta Learning released an average of six press releases per month. The difference in magnitude of the number of publications illustrates the amount of information the media produces about education compared with the provincial Ministry of Learning (McEwen, 1998).

The publishing of test scores on a regular basis has led to school, school district, and teacher comparisons. On a local level, this information has influenced the real estate market, decisions by businesses and industry, and individual decisions regarding housing and school choice (Donegan, 1998). Evidence reveals, however, that parents' choice of school is not based primarily on the school's record of high academic performance (Rubenstein, Hammer, and Adelman, 1992 in Newmann, King, and Rigdon, 1997). Educators themselves may contribute to the public's perception of the importance of test scores. Districts with high average scores publish them as evidence of their success in educating students (Nolen, Haladyna, and Haas, 1992). The Calgary Public School Board recently reversed its long standing policy of opposition to publishing school-level achievement data in the face of evidence that careful use of this information for school improvement and planning and reallocation of resources, was successfully used by the Calgary Catholic School Board to raise performance in their schools (*Calgary Herald*, January 1999). Practitioners' resistance to the release of school performance data may be fueled by blatant examples of misuse of data such as the ranking of schools, and the fear that such information is harmful to weaker schools (Raham, 1999).

Test Quality

Standardized test development is one of the most technically sophisticated specialties within education (Stake, 1998). Sound test development is a slow and expensive procedure. Instruments and operations must be examined for accuracy, relevance, and freedom from bias. A good test must have validity and reliability. A valid test is one that measures what it is intended to measure so that inferences and actions based on test scores are appropriate and accurate. If the test is reliable, the same test-taker would get the same result, regardless of who was carrying out the scoring.

In Canada, each province determines its own curriculum and ensures compliance through a central Department or Ministry of Education. Half of the provinces–British Columbia, Alberta, Saskatchewan, Quebec, Francophone New Brunswick, and Newfoundland – require students to pass centrally-set, school-leaving examinations as a condition for school graduation (Wideen, O'Shea, and Ivany, 1992). Since 1992, Manitoba has implemented language arts and mathematics examinations, and Saskatchewan and Nova Scotia each have a full complement of school leaving examinations. Ontario has just begun a secondary school literacy test which is written at the end of grade nine as a graduation requirement. Students in the Yukon write British Columbia's examinations and those in the Northwest Territories and Nunavut write Alberta's diploma examinations. Selection, accountability, and control are the underlying purposes of large-scale testing. Selection involves the sorting of individuals for future education or vocations. Accountability is designed to identify agencies within the educational system that may be held responsible for the failure or success of students–usually as groups not as individuals. Control centers on the issue of who determines what students are to learn and how that will be accomplished (Wideen, O'Shea, and Ivany, 1992). The school-leaving

examinations ostensibly serve the twin purposes of ensuring comparability across schools and reassuring the public that educational standards are being upheld.

In Alberta, the diploma examinations are a mix of multiple-choice and numerical response items and open-ended written questions. Teachers have direct input into the student grades awarded. Half of the students' final mark is based on the diploma examination score and the other half on the teacher-awarded mark based on classroom assessments. Diploma examinations are linked to standards, both for content covered and student performance, and carry high stakes for students (McEwen, 1998). Currently in Alberta diploma examinations are written for eleven core academic subjects, including English 33 and 30, Francais 30, Social Studies 33 and 30, Pure and Applied Mathematics 30, Science 30, Biology 30, Chemistry 30, and Physics 30. Five administrations of the examinations occur annually–January and June, when most students write, and August, November, and April, when selected course writings are offered. After the January and June administrations, examinations are published on the Alberta Learning Website. The annual reinvestment in test development activities has likely reduced the predictability of test content and no doubt has increased the ability of the tests to measure students' knowledge of a general domain of content and skills rather than measure how well they do on specific items familiar from previous administrations of the tests (U. S General Accounting Office, 1993).

The diplomas can be classified as standardized, criterion-referenced examinations. A standardized test, by definition, is a form of measurement that has been normed against a specific population. The items on diploma examinations have been field tested before being selected for a large-scale testing. Standardization is obtained by administering the test to a given

population and then calculating means, standard deviations, standardized scores, and percentiles. Equivalent scores are then produced for comparisons of an individual score with the norm group's performance. Criterion-referenced tests measure the achievement of specific criteria or skills in terms of absolute levels of mastery. The focus is on performance of an individual as measured against a standard or criteria rather than against performance of others who take the same test.

Diploma examinations in Alberta have three main purposes: to certify individual student achievement in specific grade 12 courses; to ensure the maintenance of province-wide standards of achievement; and to provide stakeholders with results. Alberta Learning diploma examinations are custom-built to provincial standards. Since a different test is used for each administration, efforts to minimize test score inflation is an ongoing concern. The fact that tests and test results can be misused is beyond dispute. The administration of the examinations and interpretation of results is a human endeavor. Diplomas thus are imperfect measuring devices. In Alberta, the balancing of the diploma examination mark with the teacher awarded mark for high-stake decisions such as the criteria for high school completion, supports instructional efforts that do not focus exclusively on the standardized tests. High-stakes decisions should not be based solely on single, or even multiple attempts at passing a test (Phelps, 1999).

Tests that teachers develop for their own uses are generally of very low quality, tapping into the lowest of Bloom's taxonomic levels, the category of knowledge of terms, facts, or principles (Fleming and Chambers, 1983, in Bishop, 1998a). A teacher's grades and test scores are far more likely to be idiosyncratic and non-generalizable than any standardized tests (Stiggins and Conklin, 1992). Some observers believe that without standardized tests, no one

outside the classroom can reliably gauge student progress and that standardization brings security, convenience, camaraderie, and common professional development which accompany a shared work experience (Phelps, 1999).

Provincial diploma examinations, by contrast, get a great deal of high level, professional scrutiny. Some of the most outstanding teachers in the province write the items and all items are field tested and checked for bias (Bishop, 1998b). Tests are linked to provincial curricula and measure the extent to which students have achieved provincial standards developed through the widespread involvement of teachers, provincial officials, and subject-area experts, unlike the practice in the United States where standards and assessments are not established by groups of experts who have direct responsibility for implementing curriculum and instruction. This involvement of teachers helps to increase their knowledge of curricula and instruction and facilitates the development of classroom assessments that are compatible with good instruction, and appears to have increased the acceptance of the examinations (U. S. General Accounting Office, 1993). Safeguards are designed to protect students from arbitrary test practices. Students have multiple opportunities for success, accommodations are made for those with disabilities or special needs, and a provincial funding policy tends to level resources among schools within the province in an effort to prevent gross disparities among districts and to provide students opportunities to learn the materials tested. The diploma examinations' content and technical quality has placed Alberta Learning in the forefront of large-scale assessment in North America.

## IMPROVING TEST SCORES

The purpose of this section of the literature review is to introduce the notion that it is not always desirable to improve student test scores. An exploration of the literature has unearthed a number of relevant issues regarding the enhancement of student test results. The professional development of educators in student assessment is a necessary precursor to improving student learning.

Assessment should serve one primary purpose – to improve student learning. The goal should not be to fail students, not to wave fingers at poor teachers, nor to give students more of the same even though it didn't work the first time (Kohn, 1992 ). The goal should be to provide valid and reliable information to improve student learning. Well-designed external examinations will induce improvements in instructional practice (Bishop, 1998b; Hoffman, 1999). Testing should be a tool of instruction rather than the end of instruction. Deming, a world-renowned management consultant, is often cited as the father of the quality movement in the business world–the process of continuous improvement with a clear focus on the aim or goals of the system that presumably results in extraordinary performance. Trying to manage by results, which in education means test scores, is the equivalent to driving down the road by looking in the rear-view mirror (Ramirez, 1999). Results should be used to identify which processes are most effective, and to what extent and where processes need reexamining and adjusting (Schmoker, 1996). Teachers need to be intimately involved in the development of curriculum indicators and standardized tests. We fail if we do not regularly adjust performance in light of ongoing results or if we emphasize one over the other (Wiggins, 1994). Data almost always point to action. They are the enemy of the comfortable routines.

By ignoring data, we promote inaction and inefficiency (Schmoker, 1996; Fullan, 1991; Brooks and Brooks, 1993; Glickman, 1993).

Ample evidence reveals that the climate of fear and mistrust surrounding standardized achievement tests is a national problem (Brandt, 1989, Richards, 1989, in Shepard, 1989). Increased emphasis on standardized achievement test scores by the public, the media, and schools may not have the desired effect of improvement in educational practice leading to increased scholastic achievement. Rather than raising the level of achievement, educators often seek pragmatic ways to raise test scores through means other than instructional improvement (Nolen, Haladyna and Haas, 1992). Teaching test-taking skills, promoting student motivation for the test, the development of curriculum designed to match the tests, the presentation of "practice" items similar to those used on the test, and exclusion of low-achieving students from taking tests, are among the various ways that teachers may "pollute" test scores (Nolen, Haladyna and Haas, 1992).

Teachers are constrained by fixed learning outcomes or curriculum, and the mandated external final examinations have the potential of undermining the notion of teachers as autonomous professionals. This restriction or loss of autonomy raises questions about curriculum decision making and power in education. Teachers are no longer implementors of curriculum, but are only those who deliver it (Stake, 1998). Although large-scale testing does not discourage teachers from using different approaches to learning, it has had an impact. The reduction of instructional time due to test taking activities has caused many teachers to use direct teaching rather than inquiry methods and to focus on content and algorithms required to solve set problems likely to appear on the examination rather than present a clear understanding of concepts (Samiroden, 1991). Students do not wish to undertake any activity that would

distract them from the task of preparing for the examinations. Wideen, O'Shea, and Ivany (1992) caution that the opportunity or motivation for teachers to attempt changes to their practice, a necessary condition some would posit necessary for reform, is effectively nullified by the practice of final examinations.

Misunderstandings occur when school boards pass educational goals that attempt to correlate mean scores on diploma tests to effective teaching. This over-simplification of data by an outside body needs to be challenged (Hart, 1988). Standardized achievement test scores may be used as one indicator of school or district level effectiveness. They cannot be defensively used as a measure of individual teacher effectiveness (Redfield and Craig, 1987; Raham, 1999).

The claim that all students can learn has comforted school personnel while at the same time adding even more pressure to schools where it is implied that everyone can be a rocket scientist if only teachers would teach better (Edmonds, 1984). There is a constant undercurrent of interest in evaluating how well teachers perform based on the achievement of their students (Haladyna et al., 1989; Haertel, 1986; Berk, 1988). These studies spell out in precise detail why the use of standardized test scores for a summative teacher evaluation is indefensible from both assessment and evaluation perspectives–primarily because of the insensitivity of the assessment instruments to the effects of particular teachers and the long list of factors beyond the control of the teacher that influences the scores.

> Standardized tests lack the power needed to provide in-depth information on student learning over a sufficiently long period of time and under sufficiently controlled conditions to permit them establishment of causal links between teacher performance and that of learning. (Stiggins, 1989, p.11)

Multi-indicator approaches to measuring student achievement promises to be more resistant to pollution (Nolen, Haladyna and Haas, 1992). Perhaps some of the considerable resources spent on standardized testing could be better used to improve classroom assessment and to impact schools and student learning positively.

Teacher training emphasizes the individual child, but teachers need to understand how to use data on groups of children. Data can be effective tools for promoting improvement but can never be totally accurate or reliable. In the hands of conscientious professionals, the data can be used to promote successful, goal-oriented efforts (Schmoker, 1996). To reduce the threat of misusing test result data without eliminating accountability, teachers should analyze data collaboratively and anonymously. Those closest to the point of implementation, the practitioners, need to analyze the data. Teachers should be afforded as much autonomy as possible in selecting the kind of data they think would be most helpful, and the team encouraged to exercise its own accountability (Schmoker, 1996). The number one ethical issue facing those who make curriculum policy today is where to draw the line between appropriate and inappropriate test design and implementation activities.

Need for Professional Development

School systems are under pressure to provide the public with information about what the systems are doing and how well their efforts are working. The increased pressure has led to a tightening of administrative surveillance over curriculum content, pedagogical processes in school systems, and student achievement. Teachers have been placed at the center of educational improvement efforts (Fullan and Connelly 1987, in Bosetti, 1996) and as the fulcrum, teachers must therefore be included in all stages of the assessment

process (Darling-Hammond, 1994; McEwen, 1998). The mistaken belief on the part of policy makers is that they control school quality when in fact students and teachers do.

Without serious commitment to improving teachers' skills through professional development, it is likely that the "pressure to improve test scores may well corrupt both the teaching and learning process, and the meaning of test scores" (Herman, 1997, p.6). The most important of all schooling inputs is teacher quality. Teacher experience and verbal ability are two factors that bear a stronger and more consistent relationship with student performance on standardized achievement tests than do other characteristics (King and Verstegan, 1998). In the end, one can only improve education by improving the quality of instruction–a much more complex process than designing good tests (Hoffman, 1999).

Although typical teachers can spend as much as a third to half of their professional time involved in assessment-related activities (Stiggins and Conklin, 1992), most teachers and administrators are unprepared from preservice and graduate training to assess student learning. They may lack the ability to evaluate assessment strategies because they have little if any formal assessment training (Stiggins, 1989; Stiggins, 1999a; Schafer and Lissitz, 1987). If it is desirable for schools to be accountable for student outcomes, school systems must be designed in ways that enable teachers to identify problem areas, and that provide teachers with the knowledge, power, and information needed to make meaningful changes ( Mohram et al., 1994; Kelley, 1997).

Like many large testing programs, the Learner Assessment Branch of Alberta Learning has formal reviews to scrutinize items throughout test development. Practicing teachers are involved throughout the process because of their familiarity with students and the curriculum. Their participation ensures

that the test items are appropriate for the students given the provincial expectations and Program of Study in a given subject area. These teachers are provided with basic item writing principles using content-specific manuals which contain explicit guidelines and examples, and the assistance of test development specialists (Gierl, Khaliq, and Boughton, 1999; Herman, 1999; and Newman, King, and Rigdon, 1997). Teacher involvement with Alberta Learning in diploma examination development may be the only formal assessment training they receive.

In the high standards-high involvement model of education reform, teachers assume a variety of new tasks involving school management: shared decision making for curriculum and instruction with an increased responsibility to produce high levels of student achievement (Kelley, 1997; Ramirez, 1999). Educators who have no access to useful, quality professional development will be incapable of dealing with challenging content standards, higher expectations for all students and new methods of assessment, instruction, and accountability (Watts, Gaines, and Creech, 1998). Professional development for teachers is a priority to ensure that assessment practices are understood and used to enhance student learning. Samples of school district personnel from across Canada suggest that most do not want more external testing, "they are interested in developing the kinds of links required to foster systematic change in promoting assessment strategies to maximize students learning through the use of more authentic hands-on testing, and in developing a broad range of evaluation strategies" (Mawhinney, 1998, p.108).

Teachers feel better about examinations if an effort is made to lessen confusion surrounding objectives and uses of the diploma examinations results. Ultimately, information and better understanding of policies can reduce the negative effects of examinations (Swanson, 1994). Principals and

administrators should be prepared to understand issues involving unethical and inappropriate use of assessment information and ways to protect students and staff from misuses. They have a key role in the understanding of assessment policies and regulations that contribute to the development and sound use of assessments at all levels. "It is incumbent on school leaders and policy makers to be knowledgeable about the technology of testing. Their knowledge must include an understanding of what the technology cannot do " (Ramirez, 1999). Administrators must also assume a leadership role in communicating effectively with members of the school community about assessment results and their relationship to instruction (Ramirez, 1999).

The principalship has become increasingly challenging, complex, and stressful as educational leadership continues to evolve (Davis, 1998). The metaphors of the principal have evolved from instructional leader in the 1980s, facilitator in the 1990s, to senior manager in the new millennium (Payne and Michailides, 1998). As more demand that school systems be accountable for the performance of their students, the principal is ultimately held accountable as the leader of the school. When it comes to judging the effectiveness of principals, only a few objective measures exist–mainly student achievement, drop-out rates, and standardized test scores. Principals need to provide the time and guidance for their teachers to review sample assessments and to align items with standards.

One way to deal with potential misuse of student performance data is to become assessment literate. School districts put themselves in the driver's seat when they invest in professional development and collaborative cultures that focus on student learning and the associated improvements in instructional practices (Fullan, 1998). Professional development must be aligned with standards, focused on student achievement, flexible and responsive to individual

school needs, accessible, convenient, and adequately funded (Watts, Gaines, and Creech, 1998).

A number of publications available from Alberta Learning may be instrumental in promoting assessment education and the appropriate use of diploma examination results. The *Guidelines for Interpreting and Using the Results of the Diploma Examinations* can be found in the *Annual Report: Diploma Examinations Program* and on the Alberta Learning Website. This document outlines a number of important considerations useful when interpreting diploma examination results. The guidelines encourage teachers and stakeholders to examine results in relation to provincial goals and standards. The document recognizes that diploma results provide only part of the overall picture of the province's, a school jurisdiction's or a school's performance. Many factors contribute to student achievement that a time-limited paper-and-pencil test cannot measure including many of the important learning outcomes in the Program of Studies. Among the many considerations cited in the document is a clear statement that provincial assessments shall not be used as the sole basis for evaluating teachers' performance. Since the performance of students is a result of several years in school and many other variables, diploma results cannot be solely attributed to one teacher. There are cautions about making generalizations, especially for small groups of students, and in making comparisons against provincial standards without taking local targets, contexts and plans into account. The guidelines adhere to the *Principles of Fair Student Assessment Practices for Education In Canada* (1993). An understanding of the implications of the *Access to Information (Policy 3.2.5): Right of Access to Diploma Completed Examinations Policy* (1995) document that was implemented in July 1996 in response to FOIPP, could reduce inappropriate assessment practices. As part of the openness of the testing and

reporting of diploma examinations, Alberta Learning has established a policy whereby students and parents of students under 16 may view and request a copy of the student's responses, detailed scoring results, and other scoring materials related to released examinations they have written.

Changes in teachers' obligations may be found in the *Teaching Quality Ministerial Order,* June 1997, which is a new policy that was implemented in September 1999, to replace the 1985 Teacher Evaluation Policy. The order requires teachers to develop teacher professional growth plans each year that reflect teachers' assessment of their professional learning needs, are related to the teaching quality standard, and take into account school, district, and provincial education goals and plans. Within the *Quality Teaching* document under evaluating student progress (section 3.2.4), teachers are expected to help students, parents, educators and members of the community understand the results of assessment and the implications for students.

Since standards demand collaboration and consensus, teachers need professional time and peer support on a regular basis to think and work through standards and their meaning. Everyone in the school must be familiar with the standards (Dougherty, 1998). Standard setting and marking examinations is an opportunity to share with colleagues and to develop more skill in evaluating how to evaluate their own students' work against provincial standards (Fuhrman and O'Day, 1996). High-quality, professional working conditions, especially more time to plan, to work with mentors, and to participate in professional development, could provide important incentives for teachers to perform at higher levels. These conditions can act as powerful rewards, and although they are expensive, if they are critical to school success, should be available in all schools, not allocated only to the more successful (Maeroff, 1988).

The logic for including student achievement in the teacher evaluation plan might be described as follows: student achievement is a goal of education; teachers are educators; therefore, teachers are responsible for student achievement. Further taxpayers need evidence that teachers are meeting their responsibility (Redfield, 1987). This logic assumes that teachers, rather than students, are responsible for student achievement and that all educational outcomes are measurable on standardized tests (Redfield, 1987). It should also be noted that average performance or gain is not a defensible expectation for non-average students, for example, for handicapped, disadvantaged, or gifted students (Redfield, 1987; Haertel, 1986). The problem is in determining the degree to which a student's achievement, however defined, is validly attributable to any particular source such as the teacher, program of studies, socioeconomic status, or innate ability (Millman, 1981, in Haertel, 1986; Redfield, 1987).

Cashing in on Good Scores

Serious problems arise when the pressure starts to take over and doing well on the tests becomes the only thing that matters. In parts of the United States, test scores are now the measuring stick used to judge schools (Hoffman, 1999). Student achievement, as measured by test scores, is a criterion of teacher effectiveness in one-third of all statewide teacher incentive and school incentive programs (Berk, 1988). If teachers believe that scores are used against them, they will do what is necessary to avoid low test scores (Smith, 1990). Cheating is an issue in some places in the United States where teachers are pressured to increase test scores. It is extremely rare in Alberta, the ATA believes, "perhaps because the stakes are not as high as in the US where school budgets, teachers' salaries, and even local real estate values can depend

on a school's performance on state examinations" (*Edmonton Journal*, July 16, 2000).

## AN ACCOUNTABILITY-DRIVEN REFORM

In this section of the literature review, economic incentives used to implement accountability-driven reforms are explored. The ongoing development of the 1989 Alberta incentive program provides insights into how policymakers influence and respond to public scrutiny and pressure. Teacher compensation; personnel practices tied to student achievement; and rewards and punishments, are topics related to the incentive discussion.

In Canada; test scores are not generally considered high stakes; however, in 1999 in Alberta, the government considered the introduction of its own incentive program which would have awarded performance bonuses to school boards for meeting provincially approved improvement targets. On March 11, 1999, Provincial Treasurer Stockwell Day announced the new education 2000-2001 reinvestment plan. The budget was welcomed by educators and parents as there were significant increases in funding for a number of areas including: per-student basic instruction, special needs, enrollment growth, early literacy, a teacher aide program, English as a second language enhancement, and student transportation. The controversial plan proposed by government in the reinvestment budget was the School Performance Incentive Program (SPIP). The program would provide additional, conditional, funding to school jurisdictions up to the year 2002. Jurisdictions could earn up to four per cent of their basic salary budget based on measured improvements on prescribed quantitative measures.

Teacher Compensation

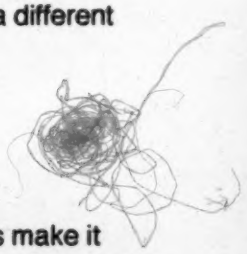
A number of teacher compensation structures have been explored in attempts to successfully implement educational polices. Traditional teacher compensation has consisted of a fixed schedule that provides salary increases for years of teaching experience and educational training or degrees. This predictable and stable practice encourages teachers to obtain advanced degrees in summer and evening courses while teaching, to attain higher levels of education and salary. No guarantee exists, however, that more years of service or education will result in more skills and knowledge. This lack of connection between educational units and school requirements has caused some to question the way we pay teachers (Knight, 1993; Mohrman and Odden, 1996).

The 1983, the *Nation at Risk* report in the US advocated merit pay. In nearly every American state where merit pay was tried, it failed. Career ladder programs have also been fraught with problems. According to Mohrman and Odden (1996), the reason for the failures is that these practices "were relatively independent of the larger educational context [and what is needed are] pay practices that enhance the core competencies on which the organization is basing its strategies" (p.52). The importance of aligning pay practices to strategic needs of the organization has been well documented (Lawler, 1990; Lawler, 1986; Schuster & Zinghein, 1992). Mohrman and Odden(1996) note that many districts made organizational changes to effectively decentralize management, such as Site-Based Management (SBM) but made no changes to the reward system to support the new roles and contributions by teachers and administrators.

Public schools are non-profit organizations in which teachers are not able to generate increased revenues through excellent performance. Odden and

Conley (1992) point out that to attract and retain teachers who can thrive in the face of increased demands of school reforms required, teacher pay has to be market competitive. As in the private sector, cost containment is an important consideration in any pay structure or bonus incentive as in school systems that are required to improve performance with stable or declining resources. The use of performance-based rewards is contingent on the willingness of the government and the public to budget such funds.

A number of states and districts have adopted School-Based Performance Award (SBPA) polices to hold professional staff and schools accountable for improvements in student performance. State programs in Kentucky, Maryland, North and South Carolina, and district programs in Texas and Colorado, utilize various reward and sanction programs to implement school reform. The Alberta SPIP proposal shared a number of features with these American programs. Although most teachers in recent studies indicate that bonuses do not act as a strong incentive for behavioral change, they are a valued reward (Heneman, 1998; Kelly and Protsik, 1997). It should be noted that programs with a continuous improvement focus tend to produce high levels of teacher stress (Heneman, 1998; Kelly, 1999) and that teachers who do not reach performance goals are more likely than others to transfer to a different school (Heneman and Milanoski, 1998).



Personnel Practices Tied to Student Achievement

There is a growing understanding that collective agreements make it next-to-impossible for school boards to use student assessment data in teacher evaluation procedures. A ground breaking, 1997 Seattle public school teacher contract altered teacher evaluation and hiring provisions. It ties teacher assignment to skills required by the school and links teacher evaluation to gains

in student achievement. In Texas, an appraisal system began in 1997-1998 that requires districts to take into account student performance when evaluating teachers. Oregon has eliminated teacher and principal tenure. Tennessee's *Value-Added Assessment System* generates confidential information on the performance of each teacher's students which principals may use in recommending professional development. The diagnostic information provided has underscored the dramatic effect skilled teachers have on performance (Quality Counts '99, 1999) and the importance of being vigilant regarding incentive programs and changes to collective agreements.

Alberta Incentive Proposal

SPIP was designed to improve student learning by encouraging jurisdictions, schools and teachers to more closely align their objectives with those outlined in the government's plan for education and in the Program of Studies. The government wanted to reward those jurisdictions, schools, and teachers who work together to improve student performance and to provide an incentive for others to continually improve (Budget 99: The Right Balance, 1999).

SPIP was a small incentive plan relative to the significant reinvestment made in education in Budget 99 and was part of the government's effort to shift "some" emphasis to the outcomes of learning. SPIP was unique when compared to other incentive schemes. The four unique components included:

1. Focus on jurisdiction rather than school level results.
2. Provisions for the inclusion of locally determined measures of improvement.
3. Absence of a penalty clause if jurisdictions participate but do not adhere to their improvement target.
4. Local board control over the bonus distribution.

The Alberta Government suggested that the bonus was not significant enough to encourage dishonesty but reminded teachers there are serious professional consequences for academic fraud. Because SPIP was jurisdiction and not individual or school based, the province suggested that it should encourage cooperation among teachers and schools and promote excellence in teaching practices to meet performance objectives. The competition was between the jurisdiction and its own past record of achievement on a number of selected performance measures. The government recognized the strength of the public education system and suggested that SPIP was not a response to a crisis, as was the case in Kentucky and other American states. SPIP was intended to recognize extra efforts made by jurisdictions to improve student learning to enhance an already strong system.

Reactions to SPIP

The response in the media was mixed but generally negative. The ATA opposes incentive programs that tie increased funding to improved student performance. In principle, they believe that funding should be invested in teaching and learning conditions, not as bonuses or rewards for improved student test scores. The ATA believed that SPIP would not contribute to improved student learning.

In the *Edmonton Journal,* Ron Chalmers (1999) briefly described the Kentucky individual school incentive plan and compared it to the SPIP– a district proposal. Because of open boundaries, the individual school plan would not work in Alberta. He suggested that principals could improve their school results by simply using selective recruitment. Chalmers said that the use of Distinguished Educators in Kentucky goes beyond Alberta's plan to help schools in which achievement may not improve. No such support initiative was

suggested for Alberta schools. Chalmers states, "Education Minister Gary Mar should take a tip from Kentucky-and from good teachers-by balancing incentives for success with remedial help for those that still are struggling" (March 11, 1999). In the same article Bauni Mackay, president of the ATA, argues that "incentives are not necessary for professional teachers".

On March 12, Mackay (*Edmonton Journal*) reminded readers that SPIP is the kind of assessment plan that was opposed by stakeholders who participated in the Alberta Government roundtables in 1993 and 1994. Don Massey, a Liberal education critic, worried that SPIP could force schools to teach to the test at the expense of other kinds of learning. He accused the government of using SPIP to wiggle out of properly funding education (*Calgary Herald*, March 12). Mackay (*Edmonton Sun*, March 12) expressed concern that SPIP was based on the assumption that teachers were not already working flat out and, "There is no evidence to suggest this is actually going to help students in their capacity to learn" (*Calgary Herald*, March 12). The chairman of the Edmonton Public School Board, George Nicholson, pointed out that his district has improved its results over the past 5 years. He was concerned that the only way for EPSB to get the incentive money was to improve on their improved level. When is enough, enough? The Grande Prairie School Council Review Committee heard from a number of parents who were also opposed to SPIP and the privatization of education. Private enterprises such as the Management Information Group promoting their "TutorSoft Achievement Software" were soliciting schools to buy their software programs to improve student overall test scores so that the jurisdiction could qualify for the new incentives. ATA vice-president, Larry Booi said SPIP, "makes as much sense as giving a thriving plant water and refusing to do the same for a dying plant until it starts doing better" (*Edmonton Journal*, May 24, 1999).

A Counter Proposal

The ATA joined three other organizations: the College of Alberta School Superintendents (CASS), The Alberta Home and School Councils' Association (AHSCA), and the Alberta School Board Association (ASBA), to formally oppose SPIP and promote an alternative incentive program. The Alberta School Improvement Program (ASIP) reflected the government's commitment to improving student achievement, however, it differed from previous initiatives in terms of the strategies to achieve this goal. In the April 28, 1999, draft document sent to the Minister, the four stakeholder groups respectfully said no to SPIP and proposed its replacement with ASIP. The draft outlined a number of problems associated with incentive programs: incentive programs do not reflect what is known about pedagogy, evaluation, practice, teacher motivation, effective school reform, or improvement practices, and noted research supporting intrinsic motivation as more important than extrinsic rewards for teachers.

In Alberta, the vast majority of teachers are committed professionals who work on behalf of students and continually seek ways to improve student achievement. The opportunity to "collaborate on school improvement efforts and to exercise more control over professional conditions of practice" is a more likely motivation for teachers than short term small economic incentives (ASIP proposal, p.2). SPIP runs the " risk of communicating to staff a perception that they are not committed to student achievement but would be enticed by incentive funding" (ASIP proposal, 1999, p.2).

There is a difference between teaching to the test and practices that improve overall student learning. Narrow quantitative measures do not reflect the total complex richness of student learning. Education research indicates that effective school improvements involve school reforms, collaborative activities

which require a local level of commitment, and appropriate funding to enable action. It should be noted that incentive programs can reward inappropriate behavior. Teachers may get the message that the only improvement outcomes are those that are financially rewarded. Standard achievement should involve both quantitative and qualitative measures over a broad range of subjects.

The highlights of the ASIP plan included:

1. Improvement programs identified by each participating jurisdiction that would reflect the local priorities as well as the priorities of the government's business plan.
2. The establishment of improvement teams for each improvement program with parents, staff and community members.
3. Improvement programs would be based on proven education research and documented pedagogical practices.
4. Program plans would be submitted to the government approximately six months before implementation.
5. Per pupil grants would be paid to fund initiatives of participating jurisdictions.
6. Identification of indicators for specific improvement plans.
7. A two year plan with documented interim reports to the government at the end of the first year and a final report after the second year.

The stakeholders recognized that, "Alberta has already established a world-class accountability system with provincial achievement tests, diploma examinations, target and goal setting activities, annual results reports and other approaches" (ASIP proposal, 1999, p.4). The ASIP proposal was an attempt to provide the government and education partners with the means for targeting areas for improvements to teaching and learning and communicating results.

SPIP– A Closer Look

Many American legal decisions, policy changes, curriculum innovations, and research findings have been assumed to bear direct relevance to the Alberta experience. SPIP assumed that education staff can be motivated by a desire for material success and that somehow if the incentive is great enough, they can readily increase their productivity (the quality of their product) through increased effort. The diploma examination program employs classroom teachers for item development and for the marking of written questions. SPIP could have reduced the teacher cooperation Alberta Learning has relied on so heavily in past examination administrations. The use of secure exams without teacher input or the use of only machine scorable questions has been discussed but is, in itself, problematic. Since current test design procedures are not intended to produce equivalent examinations it could be argued that diplomas and achievement tests give a limited measure of important jurisdiction achievement. SPIP assumed that achievement can be accurately measured and quantified as a direct result of the jurisdictions' efforts in a short period of time. It should be noted that achievement tests and diploma examinations are not perfect measures. Diploma examinations commonly fluctuate by a per cent or more between administrations as exams are built to specific blueprints but contain unique test items on each test. Thus tests are not exactly parallel to each other.

Because of the limited expertise in research design and statistics in most jurisdictions, Alberta Learning would have had to provide technical assessment support. Adding infrastructure, or more levels of bureaucracy to the education process, reduces the potential classroom dollars for student learning. SPIP also made the assumption that students are essentially interchangeable. Comparisons of the results of one cohort of students with the results of other

cohorts of students, over a short period of time, was assumed to be valid and reliable. The size of the cohort can make statistical analysis invalid, as small changes in a few students' test scores have a greater impact on a small cohort average than on the collective group test average. The effects of random variations in ability from year-cohort to year-cohort will be greater with smaller numbers.

New schools posed another problem for SPIP in that they have no baseline for student performance. Open school boundaries which allow for the recruitment of students, and other demographic changes like the introduction of new programs, are also problematic. Making comparisons from one year to the next may have little to do with the efforts of teachers or administrators.

School-level personnel focus attention on the needs of low performing students whereas ratings and rewards consider the performance of student groups. The unintended consequences of rewards and sanctions or the withholding of incentives include: the narrowing of curriculum; the altering of instructional practices to respond to priorities within accountability systems; an increase in unethical or illegal practices as the stakes are raised; the creation of morale problems and divisions among personnel as the stakes are raised; and the weeding out of the poorest students to show gain (King and Mathers, 1997).

A school's performance can only be judged fairly by taking into account many aspects of its work and many factors outside its control that affect its work.

> The attainment of the children when they first enter the school is the single most important determinant of subsequent achievement, and a growing literature exists both documenting this and discussing how fair and valid school comparisons can be made by taking it into account. (Goldstein and Cuttance, 1988, p.197)

The results of the Kentucky experience with collective incentives sheds some light on how difficult it is to design a fair and workable system of rewards based on student achievement. Alberta educators need to be vigilant in

preserving the integrity of the well run public schools in the province. The economic rewards gained by teachers in Kentucky have parents questioning why teachers are getting the money their children have earned. The hard feelings rewards and sanctions have produced in Kentucky, and the potential for similar consequences in Alberta, should be kept in mind.

Accountability-driven reforms are a very political and visible part of education reform in all provinces in Canada. They are likely to continue into the next millennium to ensure schools' accountability to society for the large public investments made in them.

Since the introduction of the SPIP concept, a number of significant changes have transpired in Alberta. Ralph Klein has reorganized his government by creating a number of new ministries with new Ministers. Alberta Education became Alberta Learning and the new Minister of Learning, Dr. Lyle Oberg, has an expanded portfolio to include Advanced Learning. A number of former aspects of the education Minister's portfolio have also been reallocated with the creation of the new ministries: Children's Services, Human Resources and Employment, and Infrastructure. In December 2000, the Learner Assessment Branch (LAB) of Alberta Learning was moved from the Basic Learning Division to the Student Information and Reporting System (SIRS) placing curriculum and student assessment in separate government departments.

The final result of the SPIP and ASIP was the implementation of the Alberta Initiative for School Improvement (AISI) proposal in October, 1999. AISI is an extension of Alberta's accountability framework. The plan encourages schools and school districts to draw on solid educational research to develop plans to meet local and provincial improvement goals that enhance student learning and results in improvements for all students. The goal is to improve

student learning and performance by fostering initiatives which reflect the unique needs and circumstances within school jurisdictions. Over the next three years, the Alberta government has pledged economic support for AISI. Unlike SPIP, funding will not be paid on bonuses; rather, it will be directed to school jurisdictions and charter schools based upon approved proposals. Acceptable proposals may be multi-year endeavors with annual progress measurement and continued funding linked to evidence of success. One of the key considerations for proposals is that they should reflect meaningful involvement of the school community. Success will be determined using a balance of qualitative and quantitative measures with a heavier weighting on local than provincial measures. The initiative results will be shared with other Alberta school jurisdictions and Alberta Learning will act as the repository on behalf of all stakeholders.

Rewards and Punishment

An educational philosophy may be based on dubious assumptions about human motivation, learning, and public policy (Kohn, 1998b). References to 'standards' or 'key performance indicators' (KPIs) tend to suggest that the primary point of schooling is to raise students' performance with respect to a very specific set of expectations for what is taught in a few basic subjects, and that this is best done by efficiently transmitting a body of facts or skills to them. The procedures for doing so can be prescribed for and imposed on teachers, and can be frequently measured by the results of standardized tests. The test scores are considered a valid basis for rewarding or punishing students, teachers, or administrators and a means to ensure that performance improves.

Upon closer inspection, raising students' performance is amazingly complicated. Too much emphasis on performance has been associated with

disturbing results: students may come to attribute their success or failure to innate ability as opposed to effort, they may have a reduced interest in learning, they may become debilitated by failure, and they may chose to avoid challenge whenever possible (Kohn, 1998). Performance can be regarded as a by-product of motivation. "Results-oriented standards do not suggest a commitment to excellence. They suggest a commitment to an outmoded, top-down approach to controlling production that is reminiscent of Frederick Taylor's scientific management 'model for factories'" (Kohn, 1998a, p.195).

Eisner believes that, "standards and performance distract us from paying attention to the building of a culture of schooling that is genuinely intellectual in character, that values questions and ideas at least as much as getting the right answer" and that the "aim of education is not to train an army that marches to the same drummer, at the same pace, toward the same destination. Such an aim may be appropriate for totalitarian societies, but it is incompatible with democratic ideals" (in Kohn, 1998a, p.195).

When people's basic need for autonomy is violated, the costs are usually high in psychological terms, and ultimately with respect to effectiveness. This is as true in classrooms as it is in countries. "When teachers are held strictly 'accountable' for their students' performance on tests – when they feel pressured to produce results– they in turn tend to pressure their students and remove opportunities for these students to direct their own learning" (Kohn, 1998, p.198). Research has shown that students are more likely to learn successfully in classrooms where the teachers had simply been invited to "facilitate" the children's learning (Kohn, 1998a; Butler and Nisan, 1986; Grolnick and Ryan, 1987).

The "prosaic mentality" (Kohn, 1998b) is a preoccupation with what can be seen and measured with the belief that any aspect of learning or life that

resists being reduced to numbers is regarded as vaguely suspicious. Concepts such as intrinsic motivation and intellectual exploration are difficult for the prosaic mind to grasp as opposed to test scores used to define success and failure. Efficiency is easier to measure than effectiveness. It is easier to determine success using numbers than to determine if what we are doing makes sense.

Research and experience demonstrate that rewards are just "control through seduction" (Kohn, 1998b). It is not surprising that when business leaders talk about getting educators to do this or that, they rely on devices described as "incentives and disincentives - in plainer words, bribes and threats" (Kohn, 1998a). Budgets can be used as instruments to garner compliance. The use of sanctions creates a climate of fear, and fear generates anger and resentment. People react by switching to damage–control mode and acting more cautiously. They stop thinking creatively and reaching for excellence when they feel they are under threat. For teachers, the result may be to become demoralized rather than motivated, accompanied by the tendency to avoid risks and play it safe. A reward-driven individual is inclined to minimize risk (Kohn, 1998b).

Factory-like schools use words like performance and achievement but rarely use words like exploration, curiosity, and discovery. Boardroom use of the concept of teamwork is usually situated in the context of competitiveness. Although social skills are often listed as desirable attributes by the private sector, they never seem to mention generosity and compassion as desirable qualities. Workers make things for profit. Kohn (1998a) states, "the only thing students should be making is meaning" (p.219).

## SUMMARY

This chapter provides a review of the literature and research on the education reforms responsible for the implementation and continuation of large-scale assessment. The literature review is not definitive given the plethora of assessment information available from a variety of sources. This review does, however, contain the few Alberta studies specifically on the use of diploma examination results.

Notwithstanding the various opinions on large-scale testing issues, it appears that there is widespread support for the diploma examination program in Alberta. In order to appreciate the impact of examinations on the behavior of students, teachers, and administrators, ongoing monitoring is necessary. The use of test scores for purposes for which they were not designed, such as school ranking, and the potential for implementation of economic incentives tied to student scores, are considered by many as unethical practices and thought to have a deleterious effect on student learning. The key to using test results effectively was intimately linked to professional development; a conclusion pervasive in the literature. It is imperative that educators and policymakers are well informed regarding the use, and potential for abuse, of examination results.

How diploma examination results are used by Alberta teachers is the purpose of this study. The relationship between the perceptions of various educational groups on the role of the diploma examinations, how well informed educators are regarding the appropriate use of student test results, and the extent that student results are used to endorse teacher competency, are research questions arising from the literature and from the 1993 Loerke study.

## 1993 DIPLOMA EXAMINATION STUDY

The Loerke (1993) study to determine if administrators used student performance on diploma examinations as a measure of teacher effectiveness and teachers' perceptions about the use of this student achievement data to evaluate them, has resulted in a number of conclusions that have implications for future study. The conclusions drawn from the findings of the 1993 Loerke study indicate that the majority of teachers were unfamiliar with the Alberta Education published guidelines for interpreting diploma examination results. This lack of awareness poses a concern as to whether these teachers interpreted student results correctly or if they analyzed the results at all. According to Loerke (1994), these teachers may not be adequately informed to discuss their students' results with various stakeholder groups, such as school-based administrators, who in this study appeared to be much better informed than teachers regarding the published guidelines and the appropriate use of student results. His findings suggest that a number of administrators and teachers involved with diploma examinations are unaware of the guidelines which may be a contributing factor to the misuse of student results in teacher assessment. Loerke (1994) contends that informed individuals are less likely to use student results in teacher assessment which is supported by the literature (Nolen, Haladyna, and Haas, 1992).

The practice of using student results in the assessment of teachers by school administrators was reported at a somewhat low level (1.4%) in the Loerke (1993) survey. The use of student results in informal teacher assessment was much more prevalent than the use of results in formal teacher assessment. A number of administrators (40%) reported that they knew or believed that student results were being used informally to assess teachers, and approximately one third of the teachers surveyed had the same perception. In his review of the

literature, Loerke (1993) suggests that a correlation between student results and teacher effectiveness has questionable reliability and that a number of other variables that contribute to student success on examinations are not under teacher control. The formal use of student results in teacher assessment can be appealed through various appeal channels. Diploma examination results that are used informally in teacher assessment present a much more serious concern because no recourse exists for such appeals.

The study also identifies differences in teacher stress associated with teaching diploma courses over non-diploma courses. The majority of survey respondents believed that teaching an externally-examined course is more stressful. The source of the stress may be self induced as teachers attempt to do the best job possible for their students, or it may be due to external forces such as perceived pressure from administrators. The size of the school appears to be another factor in how results are used and the corresponding stress felt by diploma teachers. Teachers in small schools generally reported less stress than those in medium or large-sized schools. The stress difference may be attributed to a greater certainty in how results were used. Many teachers in large-sized schools believe that results were used to varying degrees in teacher assessment. The Loerke (1993) study provides a benchmark for perceptions concerning the extent of student achievement data used in teacher evaluation. His study also leaves a number of questions that need to be addressed. The literature suggests several potential roles for diploma examination results, many of which are beyond what the examinations were designed to measure.

The purpose of the current study is to investigate the question: How are student diploma examination results being used by Alberta teachers? Aspects relevant to the purpose of this study include changes in the perceptions of administrators and classroom teachers regarding the use of diploma results

since the 1993 Loerke study, an examination of the relationship between the perceptions of school-based administrators and the perceptions of their teaching staff regarding the role of diploma examinations, an indication of how well-informed Alberta educators are regarding the legitimate use of student test results, and the extent to which student test results are being used formally or informally as an endorsement of teacher competency.

# CHAPTER 3

# RESEARCH DESIGN AND METHOD

This chapter presents details of this study's design, study population, and methodology along with a brief description of the Loerke (1993) study.

## RESEARCH DESIGN

Gall, Borg, and Gall (1996) recommend that replication studies be designed to "extend the study to be replicated in some significant way" (p.52). The design of this study was to replicate and extend the findings of the Loerke (1993) Alberta research with the addition of questions in selected areas to assess the extent to which perceptions have changed. In addition some selected interviews were held with 10 of the teachers surveyed. These provided a detailed, contextual understanding of particular situations. The addition of teacher interviews to the questionnaire data provides information that was used to indicate the extent of the reliability of the Loerke (1993) findings. Robert Yin (1989) suggests the value in using more than one method when doing case study research because it reduces the propensity for error and bias in data collection and increases the validity of the findings. Since the questionnaire instrument and the implementation conditions of the 1993 Loerke study are consistent with this comparative study, treatment fidelity is maximized and statistically more applicable (See Appendix 5).

## LOERKE (1993) STUDY

The demographics of the original survey and design, the specific variables and how they were measured, were explicit in the Loerke (1993) project and have been replicated closely. Loerke developed his own research

questionnaire since no commercial questionnaire was available. Written permission was obtained from superintendents but only verbal agreement from principals. The questionnaire cover letter established informed consent of those who returned the survey instrument. The letter to superintendents, principals, and participants did not state if the results of Loerke's 1993 research would be made available to questionnaire respondents. Questionnaires were coded for tracking and analysis purposes and were destroyed upon completion of his study. Research participants were assured anonymity as no individual names appeared on the questionnaire or in the reporting of survey data. School-based administrators assumed responsibility for the distribution and collection of the questionnaires.

The questionnaire items were primarily of a forced-choice nature. But many of the questionnaire items were also somewhat vague and open to interpretation by respondents. The researcher's telephone number was provided on the cover of the questionnaire in the event that a respondent wanted clarification of the purpose or use of the study's results. Loerke's study was specifically on the perceived use of diploma examination results in teacher assessment– a topic that was considered too specific and potentially problematic for the primary purpose of the current study.

One problem identified in the Loerke (1993) project was the classification of administrators. His study lacked the sophistication to determine whether department heads were part of the administration structure or not. It was unclear whether respondents' were a reflection of how department heads used results or of how vice-principals/principals used results. To clarity this issue, open-ended comments on the questionnaire and the use of follow-up interviews were included in this study.

Data from the questionnaires was tabulated then graphed and statistically analyzed. Frequencies for all survey questions were determined. A one-way analysis of variance (ANOVA) was completed to determine if the observations within subgroups were statistically significant. School size was one of the subgroups studied. School position was the other subgroup studied which included administrators, diploma-subject teachers, and non-diploma teachers. Administrators who taught diploma subjects were only included in the administration classification. A number of administrators in rural and small-sized urban high schools also teach diploma subjects; therefore, the blending of subcategory data was problematic. To address this classification problem, *t* tests were used in the current study to determine statistically significant differences between administrators and teachers and between diploma-subject teachers and non-diploma teachers .

The sample population in the Loerke (1993) study was representative of an entire province. Approximately 6 000 teachers and administrators were involved in the delivery of high school programs in Alberta. The study involved a stratified random cluster of high school teaching staff from schools of varying size based on the number of students writing diploma examinations. Forty two schools participated in the study. Fourteen small, medium, and large-sized school were selected. School size was based on the number of students writing diploma examinations at each school.

The questionnaire was a single method, used on a relatively large scale, on a one time basis. There was no reference in the 1993 Loerke project to a sample or pilot survey being conducted – clearly a problematic feature of the design methodology and one that has been addressed in this study.

The original written questionnaire was given to 42 of 307 Alberta high schools (14% of the total high school professional population or certified staff)

and had a 56.6% return rate. The participation rate is significant when one considers that political polls make statistically sound generalizations from 20% participation rates. Of the returns, 94% agreed to participate and those individuals who opted out did so because they were inexperienced teachers and/or had no experience with diploma subjects.

## SELECTION OF PARTICIPANTS

The selection demographics for the replication study were as close as possible to the original study. Representation for each subgroup was critical for validity purposes. The subgroups identified using biographical data included: diploma teachers, non-diploma teachers, school-based administrators, and school size. An attempt for equitable representation was made so that the numbers of diploma teachers, administrators, and non-diploma teachers selected, reflect provincial demographics. A stratified random sampling of anonymous participants from across these Alberta was solicited.

Of the 62 school jurisdictional authorities in Alberta that offer high school programs, 21 (33.9%) were selected based on geographic location and the number and/or size of schools. Superintendents were initially contacted by electronic mail (Appendix B). Interested superintendents were sent hard copies of the study description, survey instrument, and consent form that was to be signed and returned to the researcher before individual school principals were asked to participate in this study.

Eighteen superintendents granted permission for this study, representing 29.0% of the Alberta school authorities. There are approximately 710 high schools in Alberta including religious based programs, residential schools, adult education, alternative delivery programs, and upgrading institutions. Of the 71 out of 710 (10.0%) high schools in the province selected 24 (33.9%) of the

principals denied permission for this study at their school. Many of these principals expressed concerns about an overabundance of other studies and the growing number of commitments of their staff, as well as a desire not to add to the load of very busy teachers and administrators. Of the 47 principals (66.2%) who agreed to participate, 19 were from small-sized schools, 13 from medium-sized schools, and 15 from large-sized schools.

Rate of Return

Principals were asked to identify the number of teaching staff and administrators at each of their schools. A total of 802 questionnaires were sent and 470 were returned, representing a return rate of 58.6%. Of the questionnaire surveys mailed to individual schools, 53 were returned directly to the researcher and 417 were returned by the school. Eighteen questionnaires were returned late and therefore are not included in the statistical analysis. Of the 452 questionnaires returned on time for analysis, 404 respondents were willing to participate in the study and 48 elected not to complete the survey. Of the respondents willing to participate, 34 (8.4%) had completed a similar survey in 1993, presumably the Loerke (1993) study.

Profile of Respondents

Of the 452 respondents included in the analysis, 40 were administrators (21 principals and 19 vice principals) and 412 were teachers. Of the 412 teacher respondents, 49 were department heads. Of the 362 respondents who identified, whether or not they had taught a diploma course, 247 were declared diploma subject teachers and 115 non-diploma teachers. Of the 404 respondents who completed the questionnaire, 279 (69.1%) currently teach a 30 level subject, 102 (25.2%) teach a 33 level subject, and 309 (76.5%) of

respondents had a specialization or degree in their primary teaching assignment. Of these respondents, 76.0% have taught for more than 10 years and 45.3% for more than 21 years. Approximately eight percent of the respondents in this research study also participated in the 1993 Loerke study.

## SELECTION AND PROFILE OF INTERVIEWEES

Interviewees were selected to represent various stakeholders relevant to the use of diploma examinations. The selection was based on the researcher's knowledge of the individual's teaching assignments, ATA involvement, and school position. An equal number of male and female interviewees were selected from various geographic areas in Alberta. The interview representatives included: administrators, diploma teachers, non-diploma teachers, and ATA representatives. All of the individuals selected agreed to participate in the semi-standardized interviews. Due to the limited number of interviews conducted and the need to represent the perceptions of the stakeholder groups identified, potential interview candidates were acquaintances of the researcher. The interviewees' perceptions about this research topic were not known prior to the semi-standardized interviews. This interview format allowed participants to address various aspects of the role of diploma examinations. The interviews were conducted prior to the analysis of the quantitative data.

## DATA COLLECTION

This study included the use of a pilot questionnaire and trial interviews with experienced professionals who were willing to provide feedback. Their information has not been included in this study.

Pilot Study

A pilot questionnaire was designed with new biographical data, teacher comment space, and a number of additional questions designed to address how perceptions on particular variables have changed since 1993. The results of the pilot questionnaire resulted in some modifications. Twenty-three additional questions were incorporated into the study questionnaire and the topics were in the follow-up interviews.

Questionnaire

The original questionnaire contained 23 questions, 15 of which were forced-choice format (sliding scale of choices) and 8 biographical questions. Each question in the survey was identified as a variable by code and could be answered quickly by filling in the desired circles. This study's questionnaire has 42 forced-choice format (sliding scale of choices), 10 biographical information questions, and space for written comments at the end of the survey (See Appendix A). One focus of the new questionnaire was to gauge the extent that perceptions may have changed relative to the variables identified. Most of the Loerke (1993) questions have been embedded in this study's questionnaire (Appendix 3). Every effort has been made to maintain the integrity of the original questionnaire to minimize problems with internal validity of the instrument and the analysis following data collection. To gather more detailed or explanatory information, the questionnaire format was redesigned. Space for written comments on particular variables or groups of variables was included.

One way to gather questionnaire information efficiently, was to ask principals to assume responsibility for the distribution and collection of the survey instrument with their teaching staff. The involvement of the principals occurred after an initial electronic-mail contact with an explanation of the

purpose of the study and a standard consent form to initiate the survey. Copies of the superintendent and principal request-for-participation letters, participant agreement documentation, and study information and purpose cover letters may be found in Appendix B.

Interviews

The second data collection method involved semi-standardized teacher interviews to elaborate and extend the questionnaire findings (Appendix C). The selection demographics for interviews were similar to the questionnaire but on a much smaller scale. Five of the ten interviews were conducted after approximately half of the schools returned their questionnaires. The remaining interviews were held in the summer following the return of the last questionnaires. The open-ended comments on the questionnaires, in conjunction with trends in questionnaire responses, helped shape the interview topics. The first few interviews provided valuable insights and guided the subsequent set of interviews. The semi-standardized interviews contained a number of items in common with the written questionnaire. In addition, the interviews included adaptive questions to extend the understanding of the researcher on this topic. The repetitive nature of the questions should generate data that is statistically valid for a replication study.

Interviews are somewhat more adaptable than a single questionnaire. Additional questions or explanations may be generated in light of a participant's responses. Clarification of the questions and the responses are possible with the interview method. It was possible to gain information, particularly information concerning the negative aspects of teacher evaluation, that may not have been revealed using a questionnaire. This level of information is possible if

the interviews are conducted in a professional manner and a good rapport between the interviewer and the participant is established.

Timeline

The ideal time to collect data using a questionnaire or interviews for diploma examination perceptions is a couple of weeks after school authority and diploma *Examiner's* reports from Alberta Learning are sent to schools. The reports contain a detailed analysis of the diploma examination results on specific subjects for individual students and for instructional groups. These reports are prepared and sent to schools twice each year. In September, the reports are for students who wrote examinations the previous June, and in February, for the students who wrote the January diploma examination. Student examination results, interpretations of the reports, and the use/abuse of this information, are fresh in the minds of the study participants given the proximity of the examination results. The most accurate perceptions may be solicited at this time. Loerke (1993) distributed his survey following the receipt of these reports by the schools.

Late February was also the preferred survey time as it allowed for the collection of information from a relatively constant population before the end of the school year and the teachers and administrators surveyed were likely to be at the same school until the end of June. Although the researcher started this process with hopes of placing questionnaires in the schools at this time, various consent obtaining obstacles presented themselves. The researcher had to first gain the approval at the university before conducting this study, then legally, each superintendent had to give written consent before principals at schools in their jurisdictions could be approached. A number of the large school districts require research study proposals to be approved collectively by specially

appointed superintendent representatives. This process was somewhat time consuming; however, in writing proposals and rationales the researcher gained a number of valuable insights into conducting educational research.

The timeline for the data collection for this study was approximately eight months– from the writing of the study proposal for the university ethics board approval to the last interview conducted. The cost of sampling respondents using a questionnaire over a wide geographical area was somewhat low compared to the cost and time required to collect data by conducting provincial interviews. The budget for printing and posting costs for the questionnaire and for audio tapes and transcription was paid by the researcher.

## PROTECTION OF HUMAN SUBJECTS

Numerous steps have been taken to ensure the rights and anonymity of the study participants and their schools. On the questionnaire, school codes were used by the researcher to track participants, school size, and subgroups of respondents. Some school authorities demanded that school size rather than number codes be used to track questionnaire responses. A cover letter was included for each questionnaire regarding: the right to opt out, participation permission, the use of the survey information, and the efforts that would be taken to assure anonymity and confidentiality. The right to opt out may have been exercised by participants at any time during the study. In addition to the questionnaire cover letter that indicated this option to potential participants, two formal provisions for opting out of this study existed: the superintendent/principal could elect not to have teachers or administration in the identified school authorities/schools participate in the proposed study by denying permission to the researcher, and individual teachers/administrators could choose to fully or partially complete the survey questionnaire and/or participate in the interviews.

The protection of human subjects involved in this study was respected. Participants were selected equitably using random selection techniques; a written letter of intent accompanied the questionnaire that outlined the purpose of the research being conducted; written consent was obtained from each superintendent, principal, and interview participant; individual questionnaires were coded by numbers or school size that were known only to the researcher to ensure the privacy and confidentiality of respondents; and on the promise that all codes and questionnaires will be destroyed when the research is completed.

School superintendents were first contacted by electronic mail, then by letter, regarding the nature of the research survey and to request access to a sample of schools in their jurisdiction. Consent was given by each superintendent by way of a standardized form included with the request for participation letter. Principals were then similarly contacted. Principals who agreed to participate were asked to select the time and place to administer the questionnaire to their high school staff/administration and to arrange for the return of the completed surveys in sealed envelopes to a neutral person. Individual teachers and school-based administration had the choice, outlined on the questionnaire cover letter, whether or not to participate in the study. Implied consent was indicated by the participants' completion of the questionnaire.

Questionnaire participants were anonymous. While a code was used to identify the school for follow-up purposes only, respondents' names did not appear on any of the survey documentation. Records of individual schools and superintendents were destroyed upon completion of the study. The data were reported in aggregate only and the findings were not reported on a specific school basis. Anonymity of the participants was addressed in the cover letter attached to each survey questionnaire.

Ten individuals were selected for the semi-standardized interviews. At least two people in each category were interviewed so that anonymity was further enhanced. Each person was interviewed for less than 40 minutes to provide answers to questions similar to those in the questionnaire (Appendix C). Private interviews were held at the convenience of the interviewee at a location other than their workplace, if they so chose. All interviewees signed a standard consent form and interviews were audio taped with verbal permission. Interviewees were given the option of not answering particular questions or stopping the interview at any time. Identifying records were not kept any longer than was necessary to complete this study. The researcher ensured confidentiality of the participants, in that no educators' names or references to their role or workplace are reported in the study. All notes and other records of interview sessions were destroyed following the completion of the study. Participants will be made aware of the publication of this study and have the option to request a copy of the results.

The self-administered survey questionnaire used codes for tracking purposes and therefore respects the anonymity of respondents. Specific schools may not have been anonymous during data collection but after data analysis was complete, all coded records were destroyed. Questionnaire respondents could elect to return the questionnaire directly to the researcher rather than to the designated individual at the school. Interviews, by their very nature, imply that the researcher knows the interviewee's identity. The data collected and reported did not include real names. Anonymity cannot be ensured to interview participants during data collection, but confidentiality has been ensured to participants via reporting techniques. The risks for interview participants were potentially greater than for questionnaire respondents; therefore, superintendents or principals were not told the interviewees' identities.

This study was accepted by the Faculties of Education and Extension Research Ethics Board at the University of Alberta before proceeding as well as various school-district based review committees. Every effort was made to reduce risks to participants. The risks involved may be compensated by the impact of this study. The results of the study may be instrumental in positively affecting teacher assessment practices in Alberta and may be of direct value to the study participants. The benefit for participants in this research study includes the opportunity to reflect on and to discuss acceptable teacher and student assessment practices.

## DATA ANALYSIS

Loerke (1993) compared the questionnaire variables on two subgroups: administrators vs. diploma subject teachers vs. non-diploma teachers, and small vs. medium vs. large-sized schools. These subgroups provided the analytic framework for the study. The role of the participant was determined from the biographical data in the questionnaire. The organization of data collected, statistical analysis, and subsequent interpretations employ these subgroups.

The survey data generated by the present study has been compiled and sorted into similar subgroups. These groupings have been used in the analysis of the questionnaire variables, interview data collected, and for study comparison purposes. Descriptive statistics have been used to analyze the results and compare the findings of the two studies. The mean for each common variable has been calculated then presented in tables and histograms in Chapter 4 and in Appendices 2 and 3.

Factor Solution Analysis

Questionnaire variables may be commonly grouped into categories or subgroups for analysis purposes. Given the large number of variables in this study, a factor analysis method was selected. Factor analysis is a statistical procedure for reducing a set of measured variables to a smaller number of variables that are moderately or highly correlated. The smaller set of variables are referred to as a factor. After examining a number of factor solutions using loading values and the variables' contribution to the meaning of the factor, a nine factor solution was chosen. The first factor is a cluster of the most mathematically interrelated variables. Each progressive factor has fewer variables with lower interrelated values.

The factor headings identified are descriptors of the variables that loaded on each factor. The nine factors include: staff evaluation, impact on instruction, course selection, assessment expertise, teacher stress, staffing consequences, results recognition, improving results, and the use of results. Of the 42 forced-choice format questions, 37 (88.1%) were placed into one of the nine factors based on loading values greater or equal to 0.40. The researcher has interpreted the pattern of correlation to determine the conceptual meaning of the underlying factor. The remaining five variables are examined individually in the analysis discussion. Although 404 teachers participated in the questionnaire for this study, only 377 completed all the questions used in the factor analysis.

Tables consisting of three major groupings of variables: diploma teachers versus non-diploma teacher perceptions; administrators versus teacher perceptions; and small versus medium versus large-sized school for each of the nine factor solutions have been compiled. The number of participants per factor varies depending on the number of respondents who completed the set of questionnaire items for each factor identified. To determine if the variables are

significant, a one-way analysis of variance (ANOVA) on each factor was required for school-size comparisons. This statistical procedure compares the amount of between-groups variance in individuals' scores with the amount of within-groups variance. The *t* test for multiple comparison was used to test the significance of the differences between two population means for administrator versus teacher perceptions and for diploma versus non-diploma teacher perceptions. The statistics tables in Chapter 4 include F ratios and *t* values (measures of the statistically significant differences between variables). Standard deviations have been included in the analysis although the forced choices on the questionnaire do not support this 'Likert type' scale and may result in misinformation. A detailed description of the nine factor solution and analysis is provided in Chapter 4 and a list of figures and tables are provided with the table of contents.

Qualitative Data Analysis

A compilation of the anecdotal information from open-ended comments and the interviews is also included in Chapter 4. Trends and patterns among the questionnaire variables, qualitative open-ended responses on the questionnaire, and interview information were used to determine if this study was consistent with the findings of the Loerke (1993).

The qualitative data collected during the interviews was obtained using audio tapes that were transcribed, studied, and thematized. The interviews started with biographical information questions followed by questions in four related categories: how well teachers are informed about the use of diploma examinations, school/district policy regarding diploma courses, the effect of the media, and the use of student results.

The interview data collected was organized in terms of the research study questions. A content analysis of interview data was conducted before the completion of the quantitative data analysis. Given that the interview questions were similar to the questionnaire items, the data collected was used to inform the discussion of the quantitative findings.

The questionnaire comments provided descriptive data which also supplemented the statistical data presented in this chapter. Taken globally, these comments are illustrative of the range and intensity of opinions among respondents regarding the perceived role of diploma examinations. A number of themes emerged from the qualitative data collected that align with the research study questions. These findings were placed where they were most relevant. The anonymous open-ended comments were sorted by reporting groups using the biographical data and school codes from the questionnaire. The compilation of the comment data was also completed prior to the quantitative analysis.

A content analysis of the open-ended comments made by respondents on the returned questionnaires and from the interview data conducted has been included with the statistical findings where appropriate. Both quantitative and qualitative data collected were used to address the research study questions.

## VALIDITY AND RELIABILITY

To ensure that survey items in the questionnaire and those in the semi-standardized interviews had face validity, the researcher conducted a critical inspection of the items. A number of practising administrators, teachers, and a university professor were asked to provide input. Content validity was established through a pilot test, at which time respondents were told the purpose of the instrument and asked to question any items and to suggest additions or modifications. Factor solution analysis produced groupings of significant

variables. The statistical significance of the variables was determined using ANOVA and *t* tests. The significance of these variables corresponded to Loerke's (1993) study as outlined in Chapter 4.

With regard to the treatment of qualitative data, the classifications are those of the author and designed to address the research questions of the study. Perceptions of research participants based on questionnaire comments and interviews are illustrative of the range and intensity of opinion among respondents regarding the role of diploma examinations. Some comments confirmed tendencies for the entire sample of respondents and others departed from those of the average response. In either case, the responses provided insights into the use of diploma examinations in Alberta.

Given the choice of participation, and the right to opt out at any point in th study, it is uncertain whether the participants' views are representative of Alberta high school educators. There are, however, common themes that emerge from the data collected that are consistent with the Calder(1990) and Loerke(1993) findings.

## SUMMARY

The study's design was that of survey research supplemented with individual interviews. Practising administrators and teachers critically reviewed items of the questionnaire and interviews during each phase of the instrument development in order to increase the content validity of the instrument. Using more than one data collection method may increase the validity of the findings. Data were collected by means of a single questionnaire, which was pilot tested and then administered in the spring of 2000 to approximately a 10% random-stratified sample of Alberta high schools, drawn from a population of approximately 6000 high school teachers. Four hundred and fifty two usable

questionnaires were returned for an overall return rate of 58.6%. The sample was representative of the population with regard to school size, geographic location, teacher category, and school position. Quantitative data were analyzed through a variety of statistical procedures including factor analysis, calculation of means and standard deviations, as well as *t* tests, and a one-way analysis of variance. Open-ended comments and interviews were subjected to content analysis.

# CHAPTER 4

# FINDINGS

This chapter reports statistical and qualitative findings related to the purpose of the study which was to identify the perceptions of Alberta teachers concerning the role of diploma examinations results. This general purpose gave rise to five research questions that form the headings for the study's findings.

## RESEARCH QUESTIONS

The purpose of this study was to investigate the perceptions of Alberta teachers on the role of diploma examination results and the extent to which uses of diploma results have changed since Loerke's 1993 study. This general purpose gave rise to five research questions:

1. How are diploma examination results used by Alberta teachers?
2. What is the relationship between the perceptions of various educational groups, such as school-based administrators versus teachers; diploma subject teachers versus non-diploma subject teachers; and among teachers in small, medium, and large-sized schools, regarding the role of diploma examinations?
3. How well informed are high school educators regarding the legitimate use of student test results?
4. To what extent are student test results being used as an endorsement of teacher competency?
5. To what extent have teachers' perceptions of the role of diploma examinations changed since the Loerke (1993) project?

The results of the factor analysis of the questionnaire items provided an overview of the ways in which diploma examination results were used in Alberta

high schools. Factor headings provide recognizable aspects related to the questionnaire items.

The first part of this chapter reports the statistical findings of the 42 questionnaire items related to the research study to address the first research question. Teachers' responses on the 1 to 5 perception scale for each of the questionnaire items were related to various potential roles of diploma examination results. Principal component analysis was the extraction method used to determine the factor solution. The purpose of this analysis was to explore the data for underlying patterns of relationships so that the data could be described by a smaller set of items or factors. The factor analysis provided an empirical basis for reducing the large number of variables into a set of variables with a minimum loss of information (Gall, Borg, Gall, 1996). The items were factor analyzed using Varimax rotation. Items were considered to contribute to the meaning of a factor if they loaded on a factor with a value greater than or equal to 0.40 and contributed logically to the meaning on the factor. Items that loaded for more than one factor were placed into the factor for which they had the highest loading value, providing that it also contributed to the meaning of the factor. The principal component analysis identified 11 factors using the Kaiser rule (Gorsuch, 1983) as the selection criterion, whereby an eigenvalue greater than one determined the number of factors retained for further analysis. Eigenvalues measure the amount of variation in the total sample accounted for by each factor. The Kaiser criterion is a common rule of thumb used for dropping the least important factors from the analysis. The 11 factors with their respective eigenvalues and percentage of variance are shown in Appendix 4.

Rules that are typically applied when deciding the number of factors include the criterion of interpretability, the eigenvalue criterion, and the scree test (Kim & Mueller, 1978). The scree test is based on the belief that once the last

important factor has been extracted, the eigenvalues will show a discontinuity. When the factors and the eigenvalues were plotted, a straightening of the graph line occurred between Factors 9 and 10 indicating that no more than nine factors should be extracted. After running a number of factor solution analyses and considering if the loaded items contributed to the meaning of the factors, the nine factor solution appeared to be the most workable. Factors 1 to 9 accounted for 37 (88.1%) of the 42 questionnaire items. The 37 questions and their respective loading on each of the nine factors are presented on Table 1. It should be noted that real research problems are almost always more complex than the factor analysis method assumes to be true (Kim & Mueller, 1978), therefore some interpretation is required.

Of the 42 questionnaire items, 37 loaded within the outlined parameters into one or more of the nine factors. These items were used in the analysis discussion. Five items did not load within the guidelines, but have been used in the analysis where appropriate. Respondents' perceptions were reported on a 5 point scale, where 1 = knew to be true; 2= believed to be true; 3= did not know or unsure; 4 = believed to be false; and 5 = knew to be false. Item means ranged from 1.53 to 4.27. Appendix 1 recapitulates respondents' perceptions on each of the items related to the role of diploma examination results. The means and standard deviations for each factor item, as well as the overall factor mean, are reported on Table 2. The means reported on Table 2 are based on 337 participants' responses – those respondents who answered all of the items in the 9 factor solution. The means reported in Appendix 1 are representative of the total number of study participants who answered each questionnaire item (n range from 379-403). The factor means reported on Table 2 are the numerical average of the individual item means which loaded for each factor.

**Table 1: Varimax Factor Solution for the Role of Diploma Examinations Using 9 Factors**

| Questionnaire Item (Variable) | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Factor7 | Factor 8 | Factor 9 |
|---|---|---|---|---|---|---|---|---|---|
| 24. Diploma examination results are used *informally* by this district's administration as part of teacher evaluation. (V68) | 0.803 | -0.086 | 0.046 | 0.019 | 0.092 | -0.012 | 0.096 | 0.017 | -0.068 |
| 26. Diploma examination results are used *informally* by this district's administration as part of principal evaluation. (V70) | 0.800 | -0.042 | 0.008 | 0.054 | 0.095 | -0.034 | -0.112 | -0.040 | 0.176 |
| 25. Diploma examination results are used *formally* by this district's administration as part of teacher evaluation. (V69) | 0.781 | -0.075 | 0.072 | -0.119 | -0.121 | 0.108 | 0.070 | -0.023 | 0.006 |
| 27. Diploma examination results are used *formally* by this district's administration as part of principal evaluation. (V71) | 0.765 | -0.037 | 0.004 | -0.055 | -0.050 | -0.065 | -0.136 | -0.078 | 0.249 |
| 22. Diploma examination results are used *informally* by this school's administration as part of teacher evaluation. (V66) | 0.649 | 0.033 | 0.077 | -0.009 | 0.041 | 0.135 | -0.324 | 0.181 | -0.194 |
| 23. Diploma examination results are used *formally* by this school's administration as part of teacher evaluation. (V67) | 0.645 | 0.001 | 0.079 | -0.048 | -0.364 | 0.147 | -0.162 | 0.004 | -0.118 |
| 38. The use of diploma results as part of teacher assessment has increased in the past 5 years. (V82) | 0.567 | -0.127 | 0.280 | 0.089 | 0.078 | 0.217 | -0.097 | 0.144 | -0.034 |
| 31. Teachers have been promoted to leadership roles because of good diploma results. (V75) | 0.523 | -0.040 | 0.194 | 0.036 | -0.111 | 0.467 | -0.049 | 0.150 | 0.047 |
| 37. Temporary or interim teachers with good diploma results are more likely to have their teaching contracts renewed than those with lower than provincial averages. (V81) | 0.430 | -0.087 | 0.289 | 0.057 | 0.084 | 0.423 | 0.022 | 0.255 | -0.043 |

**Table 1: Varimax Factor Solution for the Role of Diploma Examinations Using 9 Factors (Continued)**

| Questionnaire Item (Variable) | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Factor 7 | Factor 8 | Factor 9 |
|---|---|---|---|---|---|---|---|---|---|
| 29. Classroom instruction has improved as a result of the reintroduction of diploma examinations in Alberta. (V73) | -0.035 | 0.822 | 0.003 | -0.033 | -0.051 | 0.030 | 0.003 | 0.022 | 0.030 |
| 20. Mandatory provincial diploma examinations have had a positive effect on student academic performance. (V64) | -0.053 | 0.788 | -0.057 | 0.026 | 0.087 | -0.032 | -0.057 | 0.046 | 0.067 |
| 49. The implementation of diplomas has resulted in the improvement of student assessment practices by teachers. (V93) | -0.065 | 0.762 | 0.095 | 0.029 | 0.002 | -0.015 | 0.197 | -0.061 | -0.037 |
| 15. The diploma examination program has been instrumental in aligning high school curriculum in Alberta. (V59) | -0.064 | 0.624 | 0.031 | -0.004 | 0.244 | -0.004 | -0.054 | 0.206 | -0.088 |
| 36. The diploma examination program has reduced the professional autonomy of teachers. (V80) | 0.170 | -0.461 | 0.229 | -0.051 | 0.194 | 0.047 | -0.111 | 0.308 | 0.163 |
| 33. Diploma teachers are as willing as non-diploma teachers to implement new curricula and/or teaching methods. (V77) | -0.062 | 0.421 | -0.026 | 0.233 | -0.184 | -0.145 | 0.118 | -0.390 | 0.197 |

**Table 1: Varimax Factor Solution for the Role of Diploma Examinations Using 9 Factors (Continued)**

| Questionnaire Item (Variable) | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Factor 7 | Factor 8 | Factor 9 |
|---|---|---|---|---|---|---|---|---|---|
| 45. At this school, students select subjects based on their anticipated success on the diploma examinations. (V89) | 0.081 | -0.066 | 0.683 | 0.028 | 0.014 | 0.049 | -0.011 | 0.032 | 0.019 |
| 44. Teachers encourage students to select particular courses based on the student's anticipated success on the diploma. (V88) | 0.025 | -0.112 | 0.681 | 0.008 | 0.028 | 0.159 | -0.142 | 0.013 | 0.052 |
| 39. At this school, efforts to improve student learning are most often focused on ways to increase diploma results. (V83) | 0.188 | -0.004 | 0.542 | -0.090 | 0.017 | 0.015 | -0.077 | 0.255 | -0.069 |
| 40. Since the introduction of 33 level diploma examinations, there has been an increase in the percentage of students at this school taking 33 over 30 level courses. (V84) | 0.055 | 0.127 | 0.539 | -0.016 | 0.073 | 0.009 | 0.161 | -0.063 | -0.127 |
| 43. At this school the grade point entrance requirement for 30 level courses is>50%. (V87) | 0.076 | 0.040 | 0.459 | 0.104 | -0.206 | -0.075 | -0.136 | -0.339 | 0.071 |
| 11. This district's administration are aware of the published guidelines and resources available for interpreting diploma examination results. (V55) | -0.030 | -0.051 | -0.015 | 0.843 | 0.041 | -0.077 | -0.060 | -0.037 | -0.083 |
| 10. This school's administration are aware of the published guidelines and resources available for interpreting diploma examination results. (V54) | 0.014 | -0.074 | -0.035 | 0.836 | 0.108 | -0.009 | 0.114 | 0.075 | -0.029 |
| 12. Diploma teachers use the Examiners' Reports and Jurisdiction Reports to analyze diploma examination results. (V56) | -0.032 | 0.095 | 0.050 | 0.531 | 0.477 | 0.082 | 0.039 | 0.019 | 0.175 |
| 50. Teachers/administrators receive adequate professional training in the assessment of student achievement. (V94) | 0.080 | 0.286 | 0.030 | 0.410 | -0.150 | -0.189 | 0.210 | 0.020 | -0.053 |

**Table 1: Varimax Factor Solution for the Role of Diploma Examinations Using 9 Factors (Continued)**

| Questionnaire Item (Variable) | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Factor 7 | Factor 8 | Factor 9 |
|---|---|---|---|---|---|---|---|---|---|
| 19. Diploma teachers at this school are given various rewards or perks in recognition of good diploma results. (V63) | 0.184 | -0.148 | 0.027 | 0.009 | -0.693 | 0.187 | 0.008 | 0.048 | 0.111 |
| 9. I have read the Alberta Learning published guidelines for interpreting diploma examination results. (V53) | 0.007 | 0.132 | 0.014 | 0.415 | 0.524 | 0.243 | 0.108 | -0.081 | 0.090 |
| 32. There is a greater level of stress associated with teaching a diploma subject than a non-diploma subject. (V76) | 0.265 | -0.140 | 0.173 | 0.102 | 0.480 | 0.162 | 0.063 | -0.079 | 0.215 |
| 34. Teachers at this school have asked for non-diploma teaching assignments because of poor diploma results. (V78) | 0.249 | -0.057 | 0.211 | -0.039 | 0.005 | 0.626 | -0.106 | 0.076 | -0.122 |
| 30. Teachers have been transferred or have asked for transfers because of poor diploma results. (V74) | 0.465 | -0.085 | 0.234 | -0.058 | 0.021 | 0.578 | 0.042 | 0.182 | -0.105 |
| 14. The administration share the details of the school's diploma results with all teaching staff. (V58) | 0.084 | -0.068 | 0.191 | 0.305 | -0.002 | -0.480 | 0.092 | 0.180 | -0.038 |
| 18. All teachers at this school are recognized for their contribution toward desirable diploma results. (V62) | -0.134 | 0.020 | 0.064 | 0.223 | -0.063 | -0.159 | 0.663 | 0.152 | 0.177 |
| 17. The publishing of individual school results by the media has had little or no effect on the way in which diploma results are used at this school. (V61) | -0.184 | 0.050 | -0.251 | -0.010 | 0.184 | 0.004 | 0.626 | -0.151 | -0.097 |

**Table 1: Varimax Factor Solution for the Role of Diploma Examinations Using 9 Factors (Continued)**

| Questionnaire Item (Variable) | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Factor 7 | Factor 8 | Factor 9 |
|---|---|---|---|---|---|---|---|---|---|
| 19. Diploma teachers at this school are given various rewards or perks in recognition of good diploma results. (V63) | 0.184 | -0.148 | 0.027 | 0.009 | -0.693 | 0.187 | 0.008 | 0.048 | 0.111 |
| 9. I have read the Alberta Learning published guidelines for interpreting diploma examination results. (V53) | 0.007 | 0.132 | 0.014 | 0.415 | 0.524 | 0.243 | 0.108 | -0.081 | 0.090 |
| 32. There is a greater level of stress associated with teaching a diploma subject than a non-diploma subject. (V76) | 0.265 | -0.140 | 0.173 | 0.102 | 0.480 | 0.162 | 0.063 | -0.079 | 0.215 |
| 34. Teachers at this school have asked for non-diploma teaching assignments because of poor diploma results. (V78) | 0.249 | -0.057 | 0.211 | -0.039 | 0.005 | 0.626 | -0.106 | 0.076 | -0.122 |
| 30. Teachers have been transferred or have asked for transfers because of poor diploma results. (V74) | 0.465 | -0.085 | 0.234 | -0.058 | 0.021 | 0.578 | 0.042 | 0.182 | -0.105 |
| 14. The administration share the details of the school's diploma results with all teaching staff. (V58) | 0.084 | -0.068 | 0.191 | 0.305 | -0.002 | -0.480 | 0.092 | 0.180 | -0.038 |
| 18. All teachers at this school are recognized for their contribution toward desirable diploma results. (V62) | -0.134 | 0.020 | 0.064 | 0.223 | -0.063 | -0.159 | 0.663 | 0.152 | 0.177 |
| 17. The publishing of individual school results by the media has had little or no effect on the way in which diploma results are used at this school. (V61) | -0.184 | 0.050 | -0.251 | -0.010 | 0.184 | 0.004 | 0.626 | -0.151 | -0.097 |

## PERCEIVED ROLE OF DIPLOMA EXAMINATIONS

This section of Chapter 4 reports the quantitative and qualitative findings relevant to the first study question: How are diploma examination results used by Alberta teachers? Headings in this section relate to the factors identified in Chapter 3. The qualitative findings are reported at the end of this section.

### Factor 1– Staff Evaluation

Many indicators have been used to determine that students are receiving a high quality education in Alberta. Results on provincial, national, and international large-scale assessments may be a reflection of the quality of instruction students have received. This factor relates the perceptions of educators on the role of diploma examination results in teacher/principal evaluation practices. The potential misuse of student results in teacher evaluation was one of the major concerns identified by teachers in the Samiroden (1991) study. Promotion to leadership roles, and the awarding of permanent contracts to temporary/interim diploma subject teachers, fit under the broad factor heading of staff evaluation. Although a number of other variables in this study could be incorporated into this topic area, only those selected by the 9 factor solution have been included.

There are nine questionnaire items related to professional staff evaluation that positively loaded for factor 1: six of the formal and informal evaluation statements, the promotion of teachers to leadership roles, the renewal of temporary teacher contracts, and the overall increase in the use of diploma

**Table 2**

**Varimax Factor Solution for the Role of Diploma Examinations Using 9 Factors**

**Means and Standard Deviations (n=337)**

| Questionnaire Item (Variable) | Item Mean | Item S.D. | Factor Mean |
|---|---|---|---|
| Factor 1: Staff Evaluation | | | 3.15 |
| 24. Diploma examination results are used *informally* by this district's administration as part of teacher evaluation. (V68) | 3.10 | 0.83 | |
| 26. Diploma examination results are used *informally* by this district's administration as part of principal evaluation. (V70) | 3.05 | 0.76 | |
| 25. Diploma examination results are used *formally* by this district's administration as part of teacher evaluation. (V69) | 3.37 | 0.84 | |
| 27. Diploma examination results are used *formally* by this district's administration as part of principal evaluation. (V71) | 3.23 | 0.70 | |
| 22. Diploma examination results are used *informally* by this school's administration as part of teacher evaluation. (V66) | 3.21 | 1.00 | |
| 23. Diploma examination results are used *formally* by this school's administration as part of teacher evaluation. (V67) | 3.71 | 0.91 | |
| 38. The use of diploma results as part of teacher assessment has increased in the past 5 years. (V82) | 2.80 | 0.74 | |
| 31. Teachers have been promoted to leadership roles because of good diploma results. (V75) | 3.18 | 0.77 | |
| 37. Temporary or interim teachers with good diploma results are more likely to have their teaching contracts renewed than those with lower than provincial averages. (V81) | 2.74 | 0.78 | |

Scale: 1 = know to be true, 2 = believe to be true, 3 = do not know/unsure, 4 = believe to be false 5 = know to be false

Continued

**Table 2: Varimax Factor Solution for the Role of Diploma Examinations (Continued)**

| Questionnaire Item (Variable) | Item Mean | Item S.D. | Factor Mean |
|---|---|---|---|
| Factor 2: Impact on Instruction | | | 2.72 |
| 29. Classroom instruction has improved as a result of the reintroduction of diploma examinations in Alberta. (V73) | 2.98 | 0.94 | |
| 20. Mandatory provincial diploma examinations have had a positive effect on student academic performance. (V64) | 2.79 | 0.95 | |
| 49. The implementation of diplomas has resulted in the improvement of student assessment practices by teachers. (V93) | 2.93 | 1.00 | |
| 15. The diploma examination program has been instrumental in aligning high school curriculum in Alberta. (V59) | 2.26 | 0.85 | |
| 36.The diploma examination program has reduced the professional autonomy of teachers (V80) | 2.68 | 0.94 | |
| 33. Diploma teachers are as willing as non-diploma teachers to implement new curricula and/or teaching methods. (V77) | 2.69 | 1.13 | |
| Factor 3: Course Selection | | | 2.63 |
| 45. At this school, students select subjects based on their anticipated success on the diploma examinations. (V89) | 2.52 | 0.94 | |
| 44. Teachers encourage students to select particular courses based on the student's anticipated success on the diploma. (V88) | 2.48 | 1.05 | |
| 39. At this school, efforts to improve student learning are most often focused on ways to increase diploma results. (V83) | 2.53 | 1.09 | |
| 40. Since the introduction of 33 level diploma examinations, there has been an increase in the percentage of students at this school taking 33 over 30 level courses. (V84) | 2.93 | 0.99 | |
| 43. At this school the grade point entrance requirement for 30 level academic courses is greater than 50%. (V87) | 2.69 | 1.55 | |

Scale: 1 = know to be true, 2 = believe to be true, 3 = do not know/unsure, 4 = believe to be false 5 = know to be false

Continued

**Table 2: Varimax Factor Solution for the Role of Diploma Examinations (Continued)**

| Questionnaire Item (Variable) | Item Mean | Item S.D. | Factor Mean |
|---|---|---|---|
| Factor 4: Assessment Expertise | | | 1.98 |
| 11. This district's administration are aware of the published guidelines and resources available for interpreting diploma examination results. (V55) | 1.69 | 0.70 | |
| 10. This school's administration are aware of the published guidelines and resources available for interpreting diploma examination results. (V54) | 1.55 | 0.69 | |
| 12. Diploma teachers use the Examiners' Reports and Jurisdiction Reports to analyze diploma examination results. (V56) | 1.64 | 0.80 | |
| 50. Teachers/administrators receive adequate professional training in the assessment of student achievement. (V94) | 3.02 | 1.10 | |
| Factor 5: Teacher Stress | | | 2.70 |
| 9. I have read the Alberta Learning published guidelines for interpreting diploma examination results. (V53) | 2.15 | 1.47 | |
| 32. There is a greater level of stress associated with teaching a diploma subject than a non-diploma subject. (V76) | 1.69 | 0.89 | |
| 19. Diploma teachers at this school are given various rewards or perks in recognition of good diploma results. (V63) | 4.27 | 0.81 | |
| Factor 6: Staffing Consequences | | | 2.67 |
| 34. Teachers at this school have asked for non-diploma teaching assignments because of poor diploma results. (V78) | 3.29 | 0.82 | |
| 30. Teachers have been transferred or have asked for transfers because of poor diploma results. (V74) | 3.20 | 0.77 | |
| 14. The administration share the details of the school's diploma results with all teaching staff. (V58) | 1.53 | 0.98 | |

Scale: 1 = know to be true, 2 = believe to be true, 3 = do not know/unsure, 4 = believe to be false 5 = know to be false

**Table 2: Varimax Factor Solution for the Role of Diploma Examinations (Continued)**

| Questionnaire Item (Variable) | Item Mean | Item S.D. | Factor Mean |
|---|---|---|---|
| Factor 7: Results Recognition | | | 2.67 |
| 18. All teachers at this school are recognized for their contribution toward desirable diploma results. (V62) | 2.54 | 1.16 | |
| 17. The publishing of individual school results by the media has had little or no effect on the way in which diploma results are used at this school. (V61) | 2.79 | 1.17 | |
| Factor 8: Improving Results | | | 3.12 |
| 21. Teachers have been provided with additional support resources and/or mentors to improve diploma results. (V65) | 3.23 | 1.17 | |
| 41. Positive recognition is given to those teachers whose school awarded marks closely align with diploma marks. (V85) | 3.08 | 1.11 | |
| 35. Teachers at this school are encouraged to include diploma results in their professional growth plans. (V79) | 3.05 | 0.99 | |
| Factor 9: Use of Results | | | 2.73 |
| 42. Students at this school are encouraged to complete 30 level diploma courses even if they are having difficulty. (V86) | 2.95 | 1.22 | |
| 46. Diploma results are an important criteria used by students/parents when selecting a high school. (V90) | 2.50 | 0.93 | |

Scale: 1 = know to be true, 2 = believe to be true, 3 = do not know/unsure, 4 = believe to be false 5 = know to be false
Continued

results as part of teacher assessment over the past five years. The last two items enumerated also loaded positively on factor 6. The means of the items in this factor ranged from 2.74 to 3.71 indicating that respondents' perceptions varied somewhat. Questionnaire items 30 and 35 also positively loaded for this factor, but were placed, respectively, in factor 6 and factor 8, based on the magnitude of the loading values and their contribution to the meaning of the factor.

Items related to the formal use of diploma examination results in teacher or principal evaluation (variables 67, 71, and 69) had means larger than 3, placing them closer to the "believe-to-be-false" category. Of the respondents who answered the items on the formal use of student results in staff assessment, between 8.3% and 9.7% knew or believed that these practices occurred compared with 27.6% to 58.1% who did not (Appendix 1). The highest percentage was reported for the formal use of results by school administrators as part of teacher evaluation (variable 67).

The means for the use of student results in the informal assessment of professional staff (variables 66,68, and 70) were smaller compared to the corresponding use in formal staff assessment. Based on mean values closer to 3, there appears to be more uncertainty about the informal use of results in teacher and principal evaluation. Of the respondents who answered these items on the informal use of results in staff assessment, between 20.2% and 22.6% knew or believed that these practices occurred compared with 22.6% to 35.6% who did not. Overall, respondents did not believe that student examination results were used formally or informally in professional staff evaluation.

Two items in this factor had means smaller than 3 placing them closer to the "believe-to-be-true" category. Overall, respondents were uncertain if the use of results had increased in the past five years (variable 82). Of the 389

respondents who answered this question, 29.1% knew or believed that this was true and 11.3% did not. Nearly 60% of the respondents were unsure if student results as part of teacher assessment had increased in the past five years. Temporary or interim teachers with good diploma results were more likely to have their teaching contracts renewed than those whose students achieved lower than provincial averages. Of the respondents that answered this question, 33.5% knew or believed that this was true compared with 13.2% who did not. The majority of respondents (53.5%) were unsure if student results were used in contract renewal decisions. Similar results were reported for the encouragement of teachers to include diploma results in their professional growth plans (variable 79). Many respondents were uncertain if diploma examination results were used in staff evaluation. Open-ended comments on the questionnaire suggested that others believed this practice was not common at their schools or in their school district. Collectively, teachers appeared to be basically uncertain about the use of student results as part of staff assessment.

Factor 2– Impact on Instruction

The impact of mandated diploma examinations on classroom instruction is of interest to stakeholders including policy makers, school-based administrators and classroom teachers. Variables loading on this factor explored the perceived changes in classroom instruction practices used to improve student academic performance, the alignment of high school curriculum, and the willingness of teachers to implement new teaching methods. Teachers' professional autonomy was a concern raised by earlier studies including Samiroden (1991) and Stake (1998). This variable was included under the impact on instruction factor heading.

Five questionnaire items positively loaded on this factor: the improvement of classroom instruction, the positive effect of diploma examinations on student academic performance, improvements in student assessment practices by teachers, the alignment of high school curriculum across the province, and the willingness of diploma teachers to implement new curricula or teaching methods. The reduction of teachers' professional autonomy was the one item that loaded negatively on this factor. The negative loading for this factor indicates that statements were dissimilar in the way in which they were written, i.e. the diploma examinations program has "reduced" the professional autonomy of teachers was written as a negative statement compared with the five other items in this factor. They were written as more positive statements using descriptors such as "improved" and "positive effect." The difference in the intent of this variable with the others in this factor resulted in the negative loading reported. The item means for this factor ranged from 2.26 to 2.98, with an average mean of 2.72.

Of the 403 respondents, 29.1% knew or believed that classroom instruction has improved as a result of the reintroduction of the diploma examinations (variable 73) and 42.0% knew or believed that the examinations had a positive effect on student performance (variable 64). That the re-implementation of the diploma program has resulted in the improvement of student assessment practices (variable 93) was less certain. The lowest mean was reported for variable 59 which suggests that respondents generally believed that examinations have been instrumental in aligning high school curriculum in Alberta. Of the 403 respondents who answered the question about whether diploma teachers were as willing as non-diploma teachers to implement new curricula and/or teaching methods (variable 77), 52.1% knew or believed this to be true. The mean for questionnaire item 36 indicates that more respondents believed that the diploma examinations had reduced the professional autonomy

of teachers (variable 80). Of those who answered this question, 43.6% knew or believed autonomy had been reduced and only 19.6% did not. The mean of 2.68 for this item indicated that overall respondents were uncertain that teachers' ability to exercise professional discretion in the delivery of instruction within the classroom had been reduced.

Factor 3–Course Selection

The variables under this factor heading investigated the topic of course selection. The success students have on diploma examinations is a result of many contributing factors some of which are outside the direct influence of school-based educators. Course selection and admission criteria have been intimately linked to school policy. Teachers' recommendations influence student course selection. Diploma examination results are one measure used to determine if student learning has improved.

Five questionnaire items positively loaded on factor 3: the ability of students to select subjects based on their anticipated success on the diploma, students being encouraged by teachers to select particular courses, efforts to improve student learning, changes in the number of students taking general over academic level courses, and the use of grade-point entrance requirements for academic courses. Means for items in this factor ranged from 2.48 to 2.93. The overall factor mean was 2.63.

Of the 390 respondents, 55.4% knew or believed that their students selected their courses based on their anticipated success on the diploma examination (variable 89) and 61.0% knew or believed that teachers encouraged students to make course selections on this basis (variable 88). Fifty-seven and one-half percent of the respondents knew or believed that efforts at their schools to improve student learning were most often focused on ways to

improve diploma results (variable 83). The majority of respondents (57.6%) knew or believed that grade-point entrance requirements greater than the 50% Alberta Learning requirement existed at their schools (variable 87). The highest mean was reported for the changes in enrollment in 33 versus 30 level subjects, suggesting that respondents did not know or were unfamiliar with the course selection trends at their schools.

Factor 4– Assessment Expertise

Variables in this factor reflected the perceived level of assessment expertise of Alberta educators. The development and administration of diploma examinations involves the collaboration of numerous stakeholders with varying degrees of expertise in student assessment and an understanding of the appropriate use of student results. Professional assessment expertise is dependent, in part, on policy makers', administrators', and teachers' awareness of appropriate assessment practices, which includes familiarity with Alberta Learning guidelines for interpreting diploma results. Professional training was identified in the literature as a key factor in maximizing the accurate and effective use of large-scale assessment results.

The four items that positively loaded on this factor were: the awareness of school districts and school administrators of the published guidelines and resources available for interpreting diploma results (variables 55 and 54); diploma teachers' use of the *Examiner's Reports* and *Jurisdiction Reports* to analyze examination results (variable 56); and the professional training teachers and administrators receive in the assessment of student achievement (variable 94). The means for items in this factor ranged from 1.55 to 3.02, with an overall average of 1.98. Item 9 also positively loaded on this factor and is relevant both

here and in factor 5. It was placed in factor 5 because of the higher factor solution value and its potential contribution to the teacher stress factor.

Respondents knew or believed that district and school administrators were aware of the Alberta Learning publications for interpreting diploma results. Respondents also knew or believed that diploma teachers used various Alberta Learning reports to analyze examination results and 69.1% of the 379 respondents who answered item 9, indicated they had read the guidelines for interpreting diploma results. What was uncertain or unknown by most respondents was if teachers and administrators received adequate professional training in the assessment of student achievement. Of the 387 respondents who answered this question, 40.3% knew or believed that their training was adequate and the same percentage knew or believed it was not.

Factor 5–Teacher Stress

In an era of public policy focused on accountability, it is not surprising that diploma-subject teachers experience increased levels of public scrutiny. The three questionnaire items that loaded on this factor relate to sources of stress for educators: the ability to interpret results, access to support resources, and an understanding of how student examination results are used.

Of the three questionnaire items loaded on this factor, the two positively-loading items were the familiarity of respondents with the Alberta Learning guidelines for interpreting diploma results (variable 53), and the increased level of stress associated with teaching a diploma subject compared to a non-diploma subject (variable 76). The perception that the diploma teachers are given various rewards or perks in recognition of good diploma results (variable 63), loaded negatively. That this variable loaded negatively was likely due to the negative connotation of receiving rewards of perks over other staff members

based on student results. By comparison, the other two variables in this factor were positively written statements. The mean for items in this factor ranged from 1.69 to 4.27, with an overall average of 2.70. Item 12 also loaded positively for this factor, but based on its loading value and content relevancy, it was placed in factor 4.

Most respondents believed that they had read the published government guidelines for interpreting results. Of the 379 respondents who answered this question, 69.1% knew or believed that this was true. They knew or believed that greater stress was associated with teaching diploma subjects than non-diploma subjects. Of the 402 respondents who answered this question, 86.1% knew or believed this to be true. Respondents also knew or believed that diploma teachers were not given rewards or perks for good examination results. Although rewards and perks may be interpreted in a variety of ways, only 2.0% of the 388 respondents who answered this question knew or believed that this was a common practice at their schools.

Factor 6– Staffing Consequences

Teaching assignments, teacher transfers, and the sharing of examination of results are the variables involved in staffing consequences related to diploma results that were addressed in this factor. The ways in which school-based administrators share large-scale test results and the value that is placed on them, have the potential to influence how examination scores are manipulated and/or used.

Two questionnaire items, the request by teachers for non-diploma subject teaching assignments, and the transfer of teachers based on poor diploma results, loaded positively on this factor, whereas the sharing of details of the school's results with all teaching staff item, loaded negatively. The first two

variables in this factor were designed as action statements that were punitive in nature. They were staffing consequences and requests for non-diploma teaching subject assignments as a result of poor diploma results. The sharing of details of student results with all staff was considered a positive administrator attribute. Thus the negative loading indicated that there was a difference in the intent of this variable compared to the others that loaded on the factor. The item means for this factor ranged from 1.53 to 3.29, with an overall average of 2.67. Items 31 and 37 also loaded positively for factor 6, but were placed in factor 1 based on the magnitude of the loading value.

Most respondents were uncertain that teaching assignments (variable 78) and transfers (variable 74) were directly related to diploma results. Of the 402 respondents who answered item 34 (variable 78), 35.5% did not know or believe that teachers had requested a change in teaching assignment due to poor diploma results. Only 12.2% knew or believed that teachers had requested transfers based on the same reason. The mean of 1.53 for item 14 indicated that most respondents knew or believed that their school administrators shared the details of the school's diploma results with the entire teaching staff.

Factor 7–Results Recognition

Diploma results are published annually on the Alberta Learning Website and in local newspapers. The increased availability of diploma results information has raised the awareness of examination results by various stakeholders. The items loading on this factor investigated if teachers at individual schools were recognized for their contribution toward desirable diploma results.

The two items positively loaded for this factor: the recognition of teachers for their contribution toward desirable diploma results (variable 62) and the effect

of the media on how schools used diploma results (variable 61). The means for these two items, respectively, were 2.54 and 2.79, with an overall average of 2.73.

Of the 388 respondents who answered item 18, 53.9% knew or believed that recognition was given to all teachers for their contributions toward student results. The mean for variable 61 indicated that they were uncertain if the media had significantly affected the way in which their school uses student examination results.

Factor 8–Improving Results

The three questionnaire items included in this factor were used to explore ways in which diploma results were improved. The three items that loaded positively on this factor were: providing teachers with positive recognition when their school awarded marks align with diploma marks (variable 85), the provision of additional support resources and/or mentors (variable 65), and the encouragement of teachers to include diploma results in their professional growth plans (variable 79). The means for these items ranged from 3.05 to 3.23.

The overall average of 3.12 indicated that most respondents were generally unsure that these activities were common practice at their schools. Of the 388 respondents who answered this question, 29.7% knew or believed that teachers were provided with additional support resources or mentors to improve diploma results compared with 46.6% of respondents who knew or believed that this to be false. The desirability of the alignment of school awarded marks with diploma marks was somewhat controversial given teachers' questionnaire comments that suggested this practice was not pedagogically sound. The alignment of the two marks was described as a "fluke at best."

Factor 9–Use of Results

School-based budgeting and open boundaries have resulted in some aggressive marketing among Alberta high schools. The items loading on this factor investigated two questionnaire items that related to the use of diploma examination results.

Two items loaded positively on this factor: the pursuit of 30 level courses and the use of diploma marks to attract students to particular high schools. The means for these two items, respectively, were 2.95 and 2.50, and the average was 2.73.

Of the 390 respondents who answered item 42, 43.6% knew or believed that students were encouraged to complete 30 level courses even if they are having difficulty(variable 86), although most were basically uncertain whether this was the practice at their school. Respondents also believed that diploma results were an important criteria used by students and parents to select a high school (variable 90).

Five items (variables 57, 60, 73, 91, and 92) did not load on the 9 factor solution analysis. They have been included where appropriate in the findings.

Qualitative Findings

A number of those interviewed, along with many who commented on the questionnaire, identified the strength of the diploma examination program as its ability to define the limits of the Program of Studies and to clarify the curriculum. They believed that the examination allowed teachers to understand the depth and breadth expected for each subject. Research participants suggested that results established a provincial standard which became the benchmark for measuring individual students and school/district success. Study participants believed that these standards level the playing field for post secondary

admissions, quota faculty admissions, employment, and various scholarship awards. Understanding the provincial standards helped teachers to establish appropriate classroom assessment standards "There has been a push in our school to standardize examinations for all subjects because teachers mark differently which raised a fairness issue," remarked a survey respondent. An examination not only credentials students, it sorts them. Most study participants believed that having consistent standards from school-to-school and from district-to-district provided the public with a clearer picture of student achievement.

Overall, respondents believed that the "misuse" of diploma examination results had started to decline. Teachers and administrators were more aware of the expectations. The recognition that five year trends and participation rates were as important to monitor as were individual examination results was positive. Most respondents were advocates for diploma examinations because of the examinations' professional quality and value as a provincial assessment benchmark. A number of participants believed in teacher professionalism. They believed that if diploma examinations were no longer mandatory, "less focus on teaching to the test" would exist, such as the addition of more relevant activities, alternative types of lessons, field trips - things that were not normally there because of time constraints." As one respondent expressed, "teachers are teaching to the exam instead of focusing on meeting the students' real needs!"

Conversely, many research respondents concerned about the demise of the diploma program worried that, "a lot of the old problems would rear their ugly heads again such as mark inflation, teachers not covering the entire curriculum, and favoritism. Kids would start shopping for the 'easiest' teachers again." If diploma examinations ceased to exist, no other provincial measuring stick for student achievement would exist; "I think the post-secondary institutions are

more concerned than anybody . . . they do not want to implement entrance exams." Many believed that students were encouraged to work harder because of the external examinations at the end of the courses, but recognized that examinations didn't measure growth, they only measured performance on a single examination.

A number of concerns were raised about the role of diploma examinations results. For many educators, the examinations appeared to be a "catch all tool" that was definitive in measuring student achievement, responsible for ensuring teachers were covered the curriculum, providing measurements of the quality of instruction, and communicating to the public how good a job any individual school was doing. Although many believed that the examinations have improved the performance of students and the responsibility for learning of teachers, the question posed by one research participant was, "Will Alberta students receive a better education without diploma examinations?"

## PERCEPTIONS OF VARIOUS EDUCATIONAL GROUPS

This section of Chapter 4 reports the findings relevant to the second study question: What is the relationship between the perceptions of various educational groups, such as: school-based administrators versus teachers; diploma-subject teachers and non-diploma subject teachers; and among teachers in small, medium, and large-sized high schools, regarding the role of diploma examinations? The headings in this section relate to the groups identified. The findings for this research question were based on statistical data generated for each grouping studied. A cross-group summary is included at the end of this section.

Factors by School Position

Table 3 reports the statistically significant differences identified in *t* tests of factors used to study the role of diploma examinations by school position. Of the nine factors measuring the perceived role of diploma examinations, five factors were identified as having statistically significant differences (at 0.01 level) between administrator and teacher respondents. A negative *t* value indicted that the means for teachers were higher on the 1-5 scale than those of administrators. Teachers reported a high level of uncertainty regarding the use of diploma examination results in professional staff evaluation. Administrators by comparison, tend to believe that diploma examination results were not used in staff evaluation (mean of 3.66). Most school-based administrators did not appear to believe that examination results were used to evaluate teachers and principals, to promote teachers to leadership roles, or as a criterion for contract renewal. Administrators reported statistically significantly higher levels of assessment expertise (mean of 1.68 versus 2.03) and higher levels of (teacher) stress associated with diploma examination subjects (1.50 versus 1.92) than was reported by teacher respondents. Administrators were more aware than teachers of the published guidelines and resources available for interpreting diploma results and that teachers used various Alberta Learning reports to analyze their students' examination performance. The majority of administrators indicated that they knew or believed that they had received adequate professional training in the assessment of student achievement. Teachers and administrator respondents knew or believed that staffing was not dependent on diploma results; however, teachers reported less certainty in how student results were used in staff placement. Administrators reported higher levels of results recognition than did teacher respondents. Administrators knew or believed that teachers were recognized for their contributions toward desirable diploma results

**Table 3**

**T Tests of Factor Scores of Measures Acceptable in Principle for the Role of Diploma Examinations by School Position**

| | Administrators (n = 40) | | Teachers (n range = 347-361) | | |
|---|---|---|---|---|---|
| Factor | Mean | S.D. | Mean | S.D. | *t* value |
| 1 Staff Evaluation | 3.66 | 0.62 | 3.08 | 0.54 | 5.71** |
| 2 Impact on Instruction | 2.95 | 0.58 | 2.90 | 0.56 | 0.60 |
| 3 Course Selection | 2.63 | 0.77 | 2.66 | 0.69 | -0.23 |
| 4 Assessment Expertise | 1.68 | 0.38 | 2.03 | 0.60 | -5.14** |
| 5 Teacher Stress | 1.50 | 0.65 | 1.92 | 0.77 | -3.74** |
| 6 Staffing Consequences | 4.09 | 0.74 | 3.57 | 0.56 | 4.28** |
| 7 Results Recognition | 1.98 | 0.83 | 2.76 | 0.91 | -5.62** |
| 8 Improving Results | 3.16 | 0.79 | 3.14 | 0.73 | 0.20 |
| 9 Use of Results | 2.48 | 0.85 | 2.75 | 0.81 | -1.93 |

** Significant at 0.01 level.

Scale: 1 = know to be true; 2 = believe to be true; 3 = do not know/unsure; 4 = believe to be false; 5 = know to be false.

and that the media had not significantly influenced how they used student marks. Generally, administrators were more decisive in answering items for the factors with the statistically significant differences reported. The factor means for administrators tended to be more polar than teacher responses. Many of the teachers' responses were near 3 indicating uncertainty. Of the remaining four factors, no statistically significant differences were identified.

Factors by School Size

Table 4 reports the statistically significant differences identified in the one-way analysis of variance (ANOVA) of factors for the perceived role of diploma examinations classified by school size. Of the nine factors measuring the perceived role of diploma examinations, five factors were identified as having statistically significant differences at a 0.01 level. Respondents from small schools (based on the number of students writing diploma examinations) knew or believed that teachers had more diploma examination expertise than their counterparts in large schools. Moreover, teachers from small schools reported higher levels of stress associated with teaching diploma examination subjects than respondents from medium or large-sized schools. Overall, respondents did not believe that results on diploma examination results affected staffing; however, respondents from large and medium-sized schools reported less certainty than teachers from small schools. Results recognition appears to be more prevalent in small schools than in large schools. Respondents from large schools reported significantly higher levels than did respondents from medium-sized schools, for the use of the results in school selection and course completion. Generally, school size appeared to make a difference in five of the nine factors selected to study the role of diploma examinations. On the

**Table 4: One-Way Analysis of Variance of Factor Scores of Measures Acceptable in Principle for the Role of Diploma Examinations Classified by School Size**

| | Small (S) | | | Medium (M) | | | Large (L) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Factor | n | Mean | S.D. | n | Mean | S.D. | n | Mean | S.D. | F Ratio | Intergroup Difference |
| 1 Staff Evaluation | 84 | 3.14 | 0.67 | 64 | 3.10 | 0.53 | 255 | 3.15 | 0.55 | 0.18 | |
| 2 Impact on Instruction | 84 | 2.83 | 0.51 | 64 | 2.88 | 0.61 | 255 | 2.92 | 0.57 | 0.84 | |
| 3 Course Selection | 81 | 2.63 | 0.80 | 58 | 2.75 | 0.76 | 252 | 2.65 | 0.65 | 0.62 | |
| 4 Assessment Expertise | 81 | 1.77 | 0.49 | 58 | 1.93 | 0.58 | 252 | 2.07 | 0.60 | 8.46** | S<L |
| 5 Teacher Stress | 84 | 1.63 | 0.67 | 64 | 1.93 | 0.84 | 255 | 1.94 | 0.77 | 5.27** | S<M, S<L |
| 6 Staffing Consequences | 84 | 3.82 | 0.67 | 64 | 3.60 | 0.56 | 255 | 3.57 | 0.58 | 5.69** | S>M, S>L |
| 7 Results Recognition | 80 | 2.38 | 0.94 | 58 | 2.55 | 0.93 | 251 | 2.79 | 0.91 | 6.50** | S<L |
| 8 Improving Results | 84 | 3.00 | 0.96 | 64 | 3.06 | 0.70 | 255 | 3.20 | 0.66 | 2.70 | |
| 9 Use of Results | 81 | 2.76 | 0.81 | 58 | 2.99 | 0.94 | 252 | 2.63 | 0.79 | 4.71** | M>L |

** Significant at 0.01 level.
Scale: 1 = know to be true; 2 = believe to be true; 3 = do not know; 4 = believe to be false; 5 = know to be false.

remaining four of the nine factors, no statistically significant differences were identified.

Factors by Courses Taught

Table 5 reports the statistically significant differences between teacher groups identified in the *t* tests of factors for the perceived role of diploma examinations. Data from administrators and other non-teaching professional staff were grouped together and placed into either the diploma or non-diploma teacher category. Respondents who indicated in the biographical section of the questionnaire that they had previously taught a diploma course were included in the diploma teacher category. If respondents indicated that they had never taught one of these courses, they were included in the non-diploma teacher category.

Of the nine factors measuring the role of diploma examinations by teacher category, three were identified as having statistically significant differences at a 0.05 level, and two were identified as having statistically significant differences at a 0.01 level. A negative *t* value implied that the mean for non-diploma teachers was higher on the 1-5 scale than that of diploma teachers. A statistically very significant difference was reported between diploma and non-diploma teachers on the impact on instruction and in perceived teacher stress. Diploma teachers reported that the examination results had a greater impact on classroom instruction than did their non-diploma teacher counterparts. Non-diploma subject teachers appeared to be more uncertain as to the effect. Statistically significant differences between groups were reported in assessment expertise, results recognition, and improving results. Diploma teachers believed that they had greater assessment expertise than was reported by non-diploma teachers. Moreover, diploma-subject teachers reported a higher

**Table 5**

**T Tests of Factor Scores of Measures Acceptable in Principle for the Role of Diploma Examinations by Teacher Category**

| | Diploma Teachers | | | Non-Diploma Teachers | | | |
|---|---|---|---|---|---|---|---|
| Factor | n | Mean | S.D. | n | Mean | S.D. | *t* value |
| 1 Staff Evaluation | 247 | 3.12 | 0.59 | 115 | 3.18 | 0.53 | -0.84 |
| 2 Impact on Instruction | 247 | 2.82 | 0.58 | 115 | 3.04 | 0.48 | -3.74** |
| 3 Course Selection | 239 | 2.68 | 0.73 | 112 | 2.62 | 0.66 | 0.79 |
| 4 Assessment Expertise | 239 | 1.94 | 0.57 | 112 | 2.08 | 0.61 | -2.07* |
| 5 Teacher Stress | 247 | 1.63 | 0.58 | 115 | 2.42 | 0.83 | -9.25** |
| 6 Staffing Consequences | 247 | 3.62 | 0.61 | 115 | 3.66 | 0.58 | -0.57 |
| 7 Results Recognition | 237 | 2.60 | 0.89 | 112 | 2.81 | 0.91 | -2.06* |
| 8 Improving Results | 247 | 3.20 | 0.80 | 115 | 3.01 | 0.58 | 2.44* |
| 9 Use of Results | 239 | 2.68 | 0.83 | 112 | 2.75 | 0.82 | -0.65 |

* Significant at 0.05 level.

** Significant at 0.01 level.

Scale: 1 = know to be true; 2 = believe to be true; 3 = do not know/unsure; 4 = believe to be false; 5 = know to be false.

level of results recognition. Respondents who had not taught diploma courses, reported higher levels of uncertainty regarding the strategies used to improve examination results. Diploma teachers believed that the strategies used at their schools were not significantly affected by diploma examination results. Generally, then, non-diploma teachers reported less certainty than diploma teachers for each of the five, statistically different, factors related to the role of diploma examinations. The means for these factors were somewhat dependent on respondents' teaching assignment. On the remaining four of the nine factors, no statistically significant differences were identified.

Factors by School Role

Table 6 reports the statistically significant differences identified using ANOVA for the perceived role of diploma examinations by school role. Data from administrators, diploma-subject teachers, and non-diploma teacher was compared to study the interrelationships of perceptions among these educators. Administrators included in this analysis indicated, in the biographical data, that they were currently not teaching nor had ever taught a diploma course.

Of the nine factors measuring the role of diploma examinations by school role, six were identified as having statistically significant differences (at 0.01 level). A statistically significant difference was reported between administrators and diploma-subject teachers and between administrators and non-diploma teachers on staff evaluation, assessment expertise, staffing consequences, and results recognition. Compared to either group of teachers, administrators reported more certainty that results were not used in staff evaluation, that educators had received adequate assessment training, and that educators were well informed regarding diploma related materials. Administrators were less likely than diploma and non-diploma teachers to believe that poor results were

**Table 6**

**One-Way Analysis of Variance of Factor Scores of Measures Acceptable in Principle for the Role of Diploma Examinations Classified by School Role**

| | Administrators (A) n = 24 | | Diploma Teachers (D) n = 223-231 | | Non-Diploma Teachers (ND) n = 83-85 | | | |
|---|---|---|---|---|---|---|---|---|
| Factor | Mean | S.D. | Mean | S.D. | Mean | S.D. | F Ratio | Intergroup Difference |
| 1 Staff Evaluation | 3.64 | 0.65 | 3.10 | 0.58 | 3.06 | 0.44 | 11.16** | A>D, A>ND |
| 2 Impact on Instruction | 3.04 | 0.61 | 2.82 | 0.58 | 3.02 | 0.48 | 4.99** | ND>D |
| 3 Course Selection | 2.52 | .060 | 2.68 | 0.70 | 2.62 | 0.64 | 0.72 | |
| 4 Assessment Expertise | 1.68 | 0.39 | 1.96 | 0.57 | 2.12 | 0.64 | 5.73** | D>A, ND>A |
| 5 Teacher Stress | 1.44 | 0.71 | 1.63 | 0.59 | 2.62 | 0.73 | 81.53** | ND>A, ND>D |
| 6 Staffing Consequences | 4.08 | 0.70 | 3.59 | 0.61 | 3.58 | 0.51 | 7.95** | A>ND, A>D |
| 7 Results Recognition | 2.02 | 0.80 | 2.65 | 0.88 | 2.98 | 0.86 | 12.11** | A>D, A>ND |
| 8 Improving Results | 3.04 | 0.74 | 3.21 | 0.80 | 2.99 | 0.58 | 3.08* | D>ND |
| 9 Use of Results | 2.38 | 0.95 | 2.70 | 0.84 | 2.86 | 0.75 | 3.29* | A>ND |

* Significant at 0.05 level.
** Significant at 0.01 level.
Scale: 1 = know to be true; 2 = believe to be true; 3 = do not know; 4 = believe to be false; 5 = know to be false.

related to teachers' requests for transfers or for non-diploma teaching assignments. Non-diploma teachers reported very statistically significant differences in perception on teachers stress. Compared to administrators and diploma teachers, these teachers reported the highest mean for the stress factor.

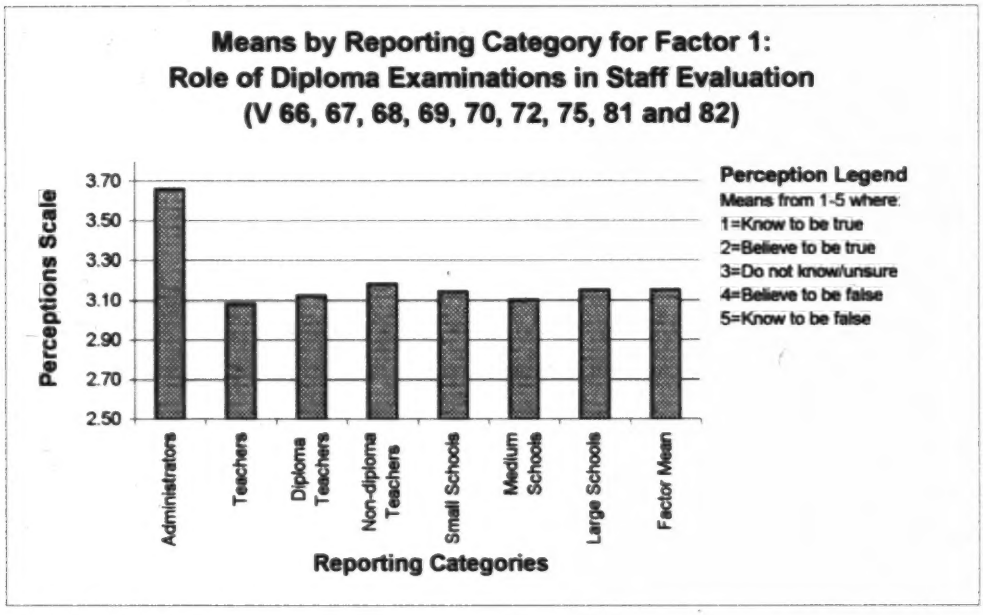
Two of the nine factors were identified as having statistically significant differences (at 0.05 level). A statistically significant difference was reported between diploma teachers and non-diploma teachers on interpreting results. Compared to non-diploma teachers, diploma teachers reported more certainty that additional resources, positive recognition, and the use of diploma results in professional growth plans were not common practices at their schools. A statistically significant difference between administrators and non-diploma teachers was reported for the use of results, whereby administrators were more likely to believe that results were not used as a criterion for selecting high schools and that students were encouraged to complete courses even when they found them difficult. Overall, administrators and diploma teachers reported similar perceptions on four of the nine factors studied: teachers stress, course selection, improving results, and use of results. No statistically significant difference was reported for Factor 3–Course Selection.

Summary of Group Perceptions

The following is a summary of the factor means and the statistically significant differences between reporting categories used to address the second research question. The means shown on the histograms are from Tables 2 through 5 in this chapter. Comparisons of means and a review of the reporting categories that were statistically different (Tables 3-5) are illustrated on the figures for each of the nine factors studied.

Figure 1 shows that among the reporting categories studied for Factor 1– Staff Evaluation, a statistically significant difference (at 0.01 level) existed between administrators and teacher respondents, only.

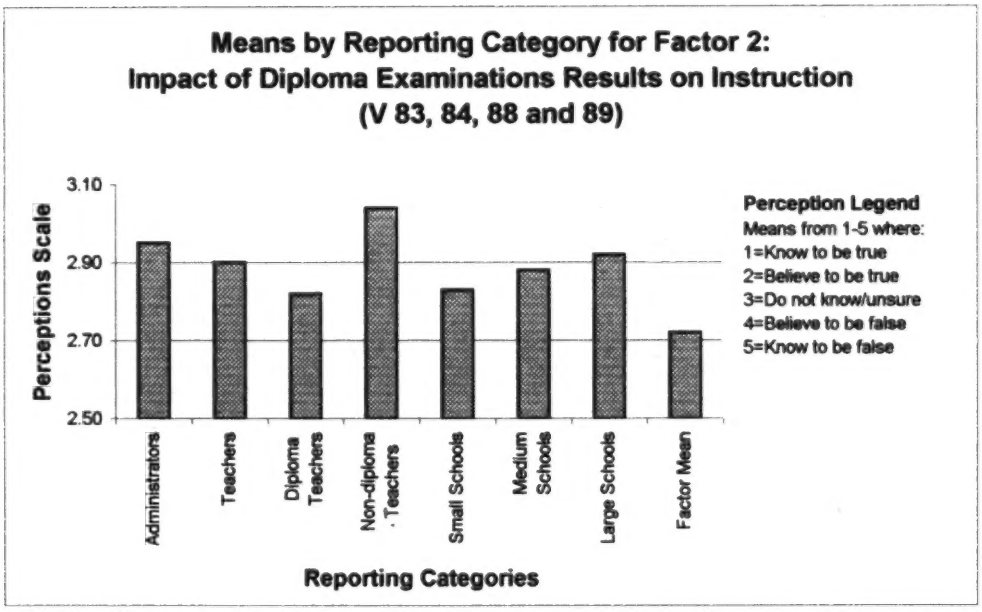
FIGURE 1




The factor mean of 3.15 was mirrored by the means for all other reporting categories, given the mean range of 3.10 to 3.18. The only significant difference in perception for this factor appears to be that of the administrators. They reported the most polar views on questionnaire items related to staff evaluation. Administrators believed that diploma examinations were less likely to be used in staff evaluation than did other respondent groups. Teacher category and school size did not appear to affect the mean for this factor.

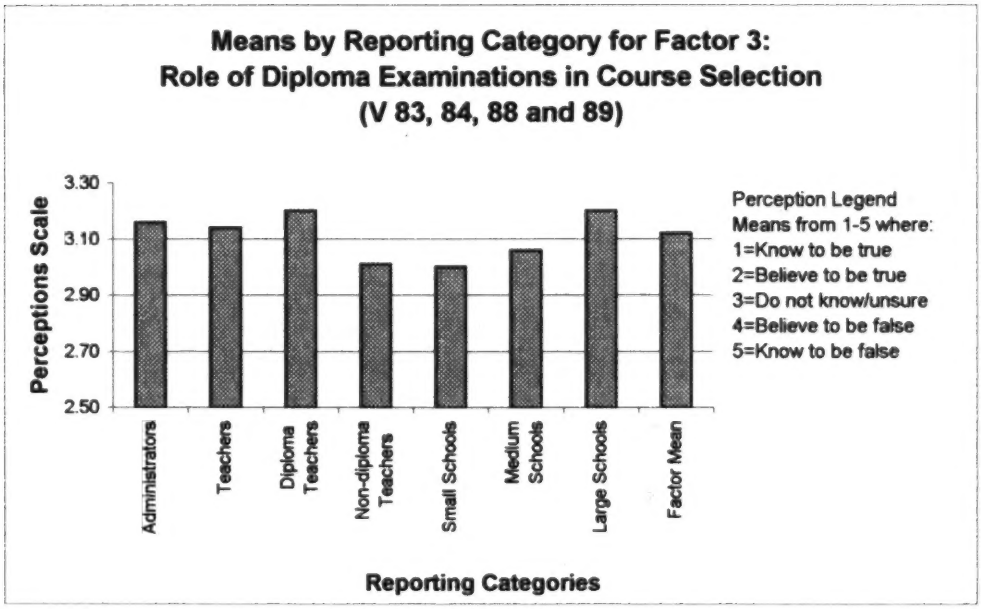
The only reporting category for Factor 2–Impact of Diploma Results on Instruction was shown on Figure 2. Statistically a significant difference was identified for diploma versus non-diploma teachers. Diploma teachers believed that examination results had greater impact on instruction than did their non-diploma teacher counterparts. School position and school size appeared to have little or no effect on the means for this factor.

FIGURE 2



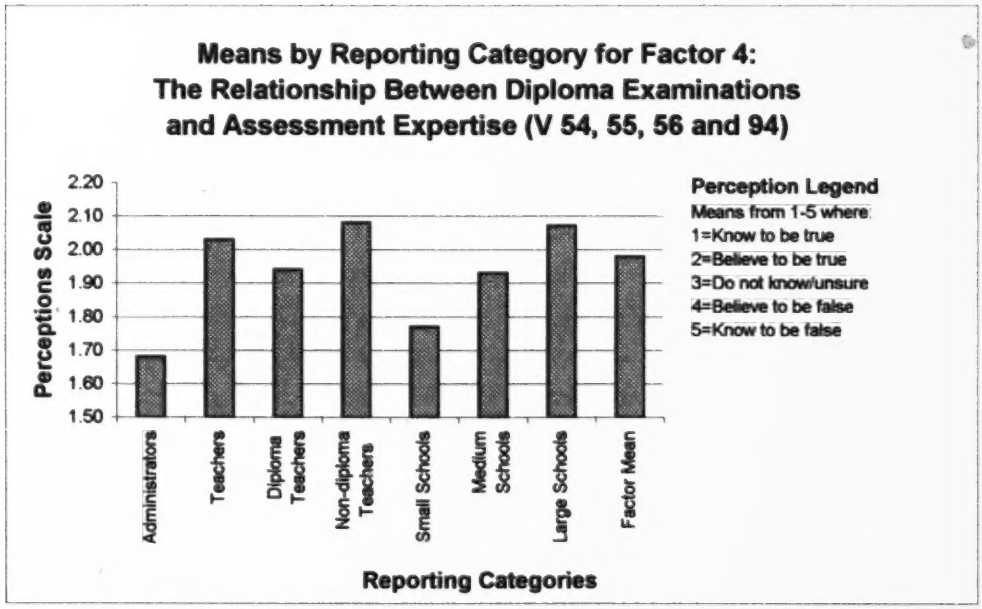
No statistically significant differences were identified in any of the reporting categories for Factor 3–Course Selection. The data shown on Figure 3 illustrate how close each of the reporting groups were. All respondents appeared to believe that diploma examinations played a role in course selection, albeit a moderate one.

FIGURE 3



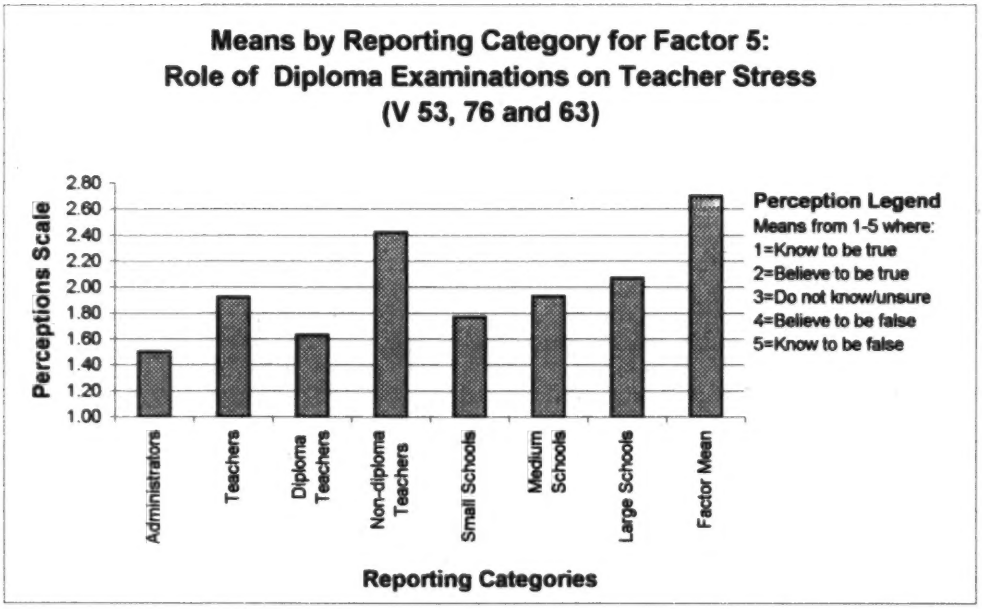
Statistically significant differences were identified in each of the selected reporting categories shown on Figure 4 for Factor 4–Assessment Expertise. Statistically significant differences were reported between administrators and teachers, and between respondents from small and large-sized schools. A statistically significant difference was identified between diploma and non-diploma subject teacher respondents. Overall, the results suggested that administrators and diploma subject teachers from small-sized had higher levels of assessment expertise. By comparison, non-diploma subject teachers and/or administrators from large schools reported lower levels of perceived assessment expertise.

FIGURE 4



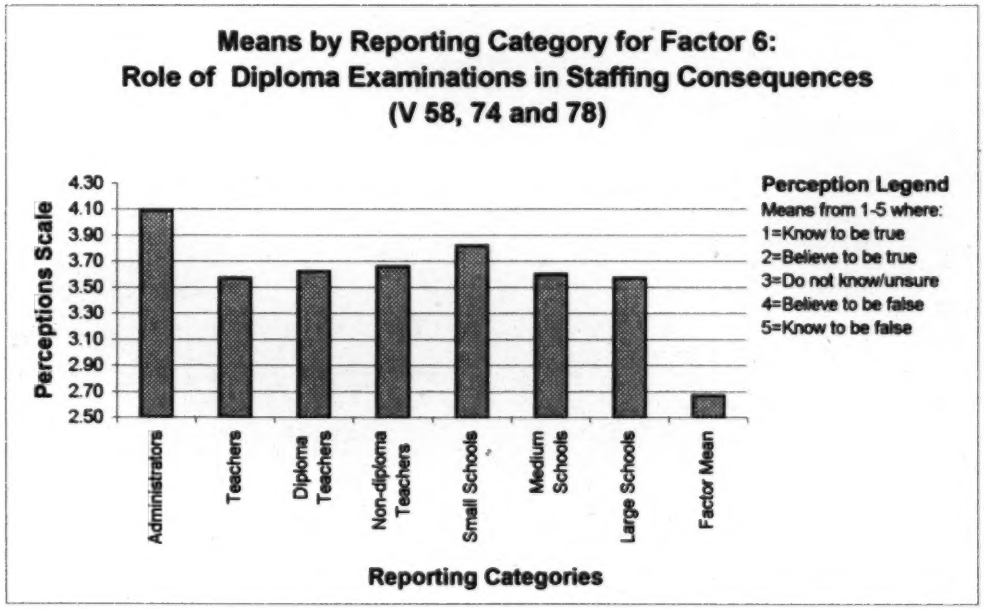
Statistically significant differences (at 0.01 level) were identified for each reporting category of Factor 5–Teacher Stress as shown on Figure 5. Collectively, the highest level of teacher stress was reported by administrators, diploma teachers, and respondents from small-sized schools. Non-diploma subject teachers and respondents from large schools reported lower levels of teacher stress.

FIGURE 5



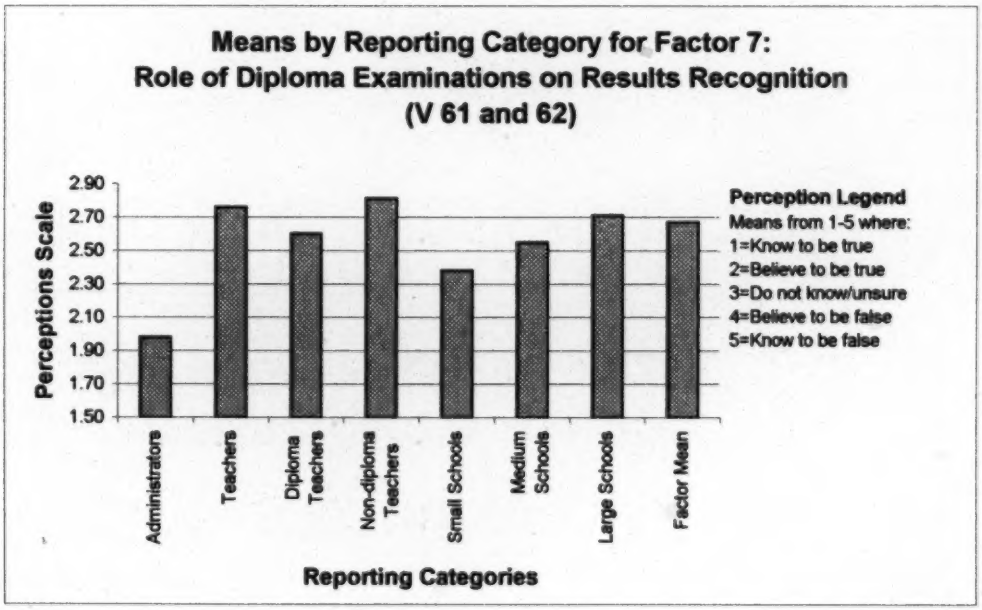
Statistically very significant differences were identified for two of the reporting categories for Factor 6–Staffing Consequences. As illustrated in Figure 6, all respondents knew or believed that the staffing consequences identified in the questionnaire were not related to student examination results. Administrators and respondents from small-sized schools reported the highest level of disbelief for this factor.

FIGURE 6



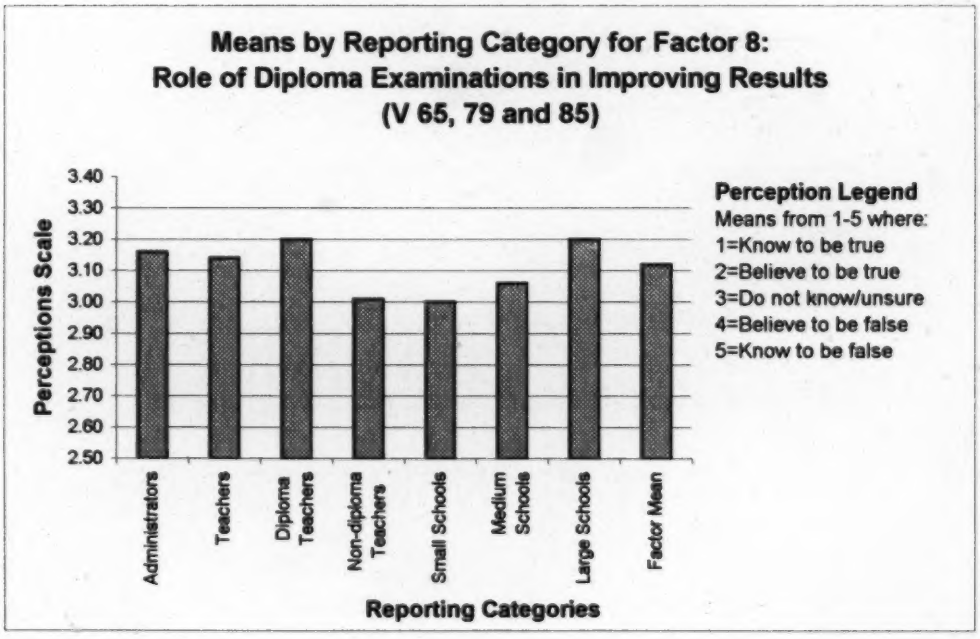
Statistically significant differences were identified in each of the reporting categories for Factor 7–Results Recognition. Statistically significant differences were reported between administrators and teachers, and between respondents from small and large-sized schools. In Figure 7, all respondents appeared to believe that results recognition was related to diploma examinations. The strongest perceptions were reported by administrators, diploma teachers, and respondents in small-sized schools.

FIGURE 7



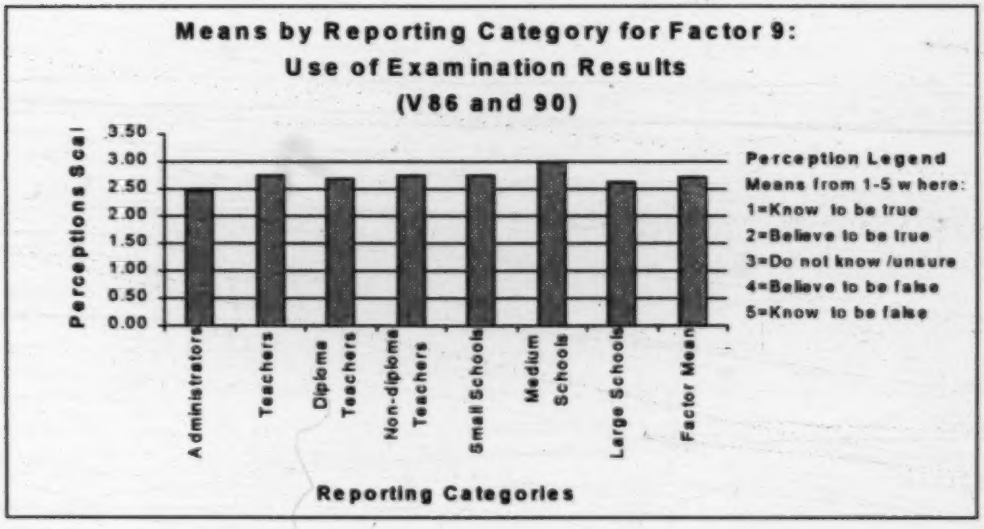
For Factor 8–Improving Results, a statistically significant difference was identified between diploma and non-diploma subject teachers, only. Non-diploma teachers appeared less certain than diploma teachers about the efforts made improve examination results. Diploma-subject teachers reported disbelief on the variables in this factor as shown in Figure 8. They knew or believed that diploma-subject teachers had not been provided with additional support resources or mentors to improve results and that teachers were not encouraged to include diploma results in their growth plans.

FIGURE 8



A statistically significant difference in Factor 9–Use of Examination Results was identified between respondents from medium and large-sized schools, only. Figure 9 shows that teachers in large-sized schools believed that diploma results were an important criterion used by the public to select high schools and that course completion was encouraged. Respondents from medium-sized schools were uncertain about the two variables in this factor.

FIGURE 9



## HOW WELL INFORMED ARE ALBERTA TEACHERS?

This section of Chapter 4 reports the findings relevant to the third research question: How well informed are high school educators regarding the legitimate use of student test results? Much of the data that follow were collected from questionnaire comments and interviews. The headings were thematic groupings that emerged from the data analysis which were relevant to how well informed high school educators were regarding the use of diploma examination results.

### Awareness of Alberta Learning Documents

Other than the non-diploma teachers interviewed, most interviewees were aware of the Alberta Learning published guidelines for interpreting diploma examination results. Most believed that school administrators and district administrators were familiar with the guidelines, various reports and resources available for interpreting results. Some participants suggested that although administrators were somewhat aware of the guidelines, time constraints caused them to focus only on diploma averages. Research participants believed that administrators "superficial look at results" did not take into account considerations such as participation rates, student ability, teacher experience, and other contextual factors.

Diploma teachers commonly used the published government information for analysis purposes. The amount of information about results provided to staff was dependent on the respondents' role in the school, and how they obtained the information varied from learning about averages in local newspapers to in-depth discussions with subject specialists. The preparation of formal, examination-result reports that were presented, in person, to central office administrators was new for many research participants. Non-diploma teachers

who responded in the open-ended comments stated that they assumed that the results for students at their school indicted that they were performing well unless administrators or the local paper made them aware that students were underachieving. Participants believed that "no news" is interpreted as "good news." Diploma teachers were most directly involved in the interpretation of students' results. Although most diploma course teachers indicated that they received a list of their students' marks and relevant reports, many were unfamiliar with the results in other subject areas.

Assessment Training

Research participants had little formal assessment training, with only a few who indicated that they had taken student assessment courses at university. Teachers appeared to rely heavily on Alberta Learning publications such as subject Bulletins to interpret the Program of Studies, and the use of school/class examination reports, along with Examiner's reports, to analyze their students' results. Teachers used these documents to ensure that they were covering the curriculum adequately. A concern was raised by a number of experienced teachers that the "exam has become the curriculum." A number of participants did not believe that teachers and administrators had time to read or possessed the "know how" to interpret the data provided, or the necessary skills to apply the information to classroom assessment practices.

Participants recognized that many of the documents from Alberta Learning are only available on the Internet. Although teachers are more technologically literate, many teachers confessed that they did not access all of the necessary information on either a regular or timely basis. Many participants suspected that students were more familiar with what is on the Alberta Learning web site than their teachers. Some teachers and administrators indicated that

they were still experiencing difficulty accessing Internet information and various documents, given limited computer availability, quality, and time.

A number of study participants believed that, ". . . we learn more as we go and aren't prepared well enough to teach diploma courses. A stronger mentorship program would help in this regard." Participants in schools with fairly transient teacher populations, believed that teachers who had previously taught junior high were hired and assigned grade 12 diploma courses, sometimes as early as their first year in high school. Research participants believed that most mentorship was informal, and dependent on their colleagues' expertise and available out-of-class time. Many participants cited teachers' involvement with the Alberta Learning field test program and the scoring of diploma examinations as valuable professional development activities.

Comments on the questionnaires indicated that participants needed more school-based information regarding specific examination administrations for particular students. *Examiner's Reports* only provided provincial feedback. More individual school and class information was desired, given the added responsibility placed on diploma teachers of writing school reports and/or explaining achieved results.

Results Sharing

Diploma results information was commonly disseminated to curriculum leaders who reviewed the data and were expected to pass it on to teachers. Both the teachers and the administrators interviewed were not sure if the follow through was adequate and expressed the need to do better follow-up of results information to all stakeholders. Participants suggested that their schools and/or districts did not have adequate funding to allow for release time of curriculum leaders and diploma teachers to facilitate the joint analysis of results, to set

learning goals, or to gain assessment expertise. Others stated that they were "given ample opportunity to participate in professional training, but chose not to."

Although many participants believed that administration and staff were given adequate information and guidelines on student evaluation, they were "not altogether sure that such guidelines are entirely appropriate. There is plenty of quantifying methodology in student assessment and not enough qualifying." More schools were focusing on five year trends in diploma results rather than on individual examination administrations. Research participants believed that the volume of assessment information that had been made available as a result of diploma examinations has helped classroom teachers to revitalize and improve their assessment skills.

School/District Policy

The grade-point entrance requirement for all high school courses set by Alberta Learning is 50%. This requirement was known by all interviewees and recognized as the legal basis for entrance into 30/33 level courses. Most schools did, however, have a "recommended entrance requirement" in place, somewhere between 60% and 65% according to questionnaire comments for item 43. Although most participants believed students were treated on an individual basis and the recommended entrance requirements were designed to ensure success in the next course level, there was a perception that these requirements also ensured higher course completion rates in grade 12 subjects and resulted in higher overall course means. Although entrance requirements set the guidelines for course admission, participants noted that in grade 12 courses, students had the right to enter courses with the minimum 50% average and allowed to write examinations even if their chances of success were questionable. Many of the interviewees believed that students were counseled

to not write examinations if they had missed "too many classes" or were "ill prepared." The use of contracts for courses for which students had lower than recommended prerequisite marks appeared to be commonplace. Many research participants believed that contracts helped students improve their work habits and attendance patterns, and that students had "the right to fail." The overwhelming belief among participants was that students and their parents made course choices and that the final decision as to whether or not to write a particular diploma examination was at their discretion.

There were a number of comments on the questionnaires about English and Social Studies 30 to 33 course transfers at, or near, the time of diploma writing. Students in a number of schools with low marks in Biology 20 and Chemistry 20 were encouraged to take Science 30, rather than Biology 30 and Chemistry 30. Many research participants felt that these practices were as much a salvation for grade-point averages for the school as they were for individual students. Some believed that "too many students are placed in 33 level courses who could handle 30 level courses but would get a lower mark. Some teachers are weeding students out of their 20 level courses instead of dealing with them [students] in their 30 level courses." Concerns were also raised that honor students in 33 level courses were not eligible for some scholarships, making the 30 level courses more desirable for the wrong reason.

Questionnaire comments indicated that students should be encouraged to complete the highest level of courses possible to ensure the greatest access to post-secondary studies, ". . . to get as much of the broad-based educational background as one is capable of getting and use this to access career choices thereafter."

The Effect of the Media

According to research participants, the media often reported student examination results "out of context" or with "no discernible context at all." A number of concerns were raised about the Fraser Institute regarding the use of diploma examination results to rank schools and the publishing of these rankings in provincial newspapers. Administrators were most vocal about the comparisons made among schools and described the situation as, "very frustrating as an administrator when you have no control over a lot of things that happen. You might not be evaluated but you are definitely studied!" Participants recognized that top ranking schools which offered specialized programs were more likely to encourage the continuation of this practice because their schools had been viewed positively by the public. Many respondents stated that high academic schools attract high academic students, regardless of media input, others believed that if diploma results were no longer published, there would be a loss of "free advertising" for the specialized schools or programs to attract high achieving students.

Although participants recognized that some newspapers are better than others in supplying a context reference for the data presented, most believed that their lives would be less stressful if such publications stopped comparing schools. "What qualifies them to make school ranking judgments?" also became a key question. A number of research participants recognized that the media were not the only source of diploma results information. They cited the posting of "superlative" subject averages and names of scholarship-winning students by school administrators and teaching staff during open-house activities as another source of "results information." Teachers and administrators appeared to be contributing to diploma results "hype."

Most respondents believed the public has a right to know how students, and, consequently schools, performed on diploma examinations, but as

educators they found it difficult, at times, to explain less than desirable results. One respondent indicated that as a parent he looked at school results on the Internet, "... if there are no students getting an 80% in academic subjects at a particular school, I'd be asking why not?" Many of the factors that contributed to the academic success of students have little to do with the quality of education provided at school. One respondent noted that, "The media commonly ignores the private and separate schools which can select their student population. Public schools, by their very nature accept all students."

Understanding Demographics

Study participants identified a number of demographic factors that affected student performance including: socio-economic status of the family, student ability, language skills, transience of the school population, immigrant population of the school, teacher ability and experience, parental relationships, students' cultural capital, and availability of specialized programs. Assertions by the media that economics and parental involvement have little effect on diploma results is not a perception shared by the educators surveyed or that of the literature studied. A number of interviewees suggested that we have work to do in educating the parents, the public, and the staff regarding the interpretation of media reports.

Ranking

Teachers in small towns had been asked "What's wrong with the school?" if the ranking of students' diploma results were below provincial standards. The writing of school reports helps address this question and has been instrumental in promoting reflective practice. The impact of the media has been mixed. Teachers and administrators have been asked better questions about how well

the courses were delivered and how to improve on existing levels of achievement compared to provincial standards over time, but often they felt forced to explain results which were not under their control.

School Selection

According to many of the study participants, the key element for high school selection continued to be where a student's friends attended school, rather than how well last year's students performed on diploma examinations or the ranking of the school in the local paper. Respondents believed that the extent to which students and their parents selected a high school on the basis of diploma examination results depended largely on the values of the parents. In some cases, parents have purchased a house in the area so that their child could attend a particular school. Respondents from small centers did not have this option as they, "are the only school in the area - there is no choice!"

## STUDENT RESULTS AS A MEASURE OF TEACHER COMPETENCY

In this section, the findings address the fourth research question: To what extent were student test results used as an endorsement of teacher competency? The qualitative data presented provide a sample of the perceptions of research participants on teacher evaluation issues as they relate to the role of diploma examinations.

The perception that "good results mean good teaching" was not necessarily a view shared by many interviewees and those who commented on the questionnaire. Many respondents agreed with one of the participants who suggested that, "it is an injustice to the teachers who are criticized when results are not as good as someone thinks they should be or praised for results which

are not really good given the ability of the group of students." Many respondents supported the view of another participant who stated that,

> too much time and effort is spent 'decoding' the media's 'oversimplified' message which has the potential of promoting an unhealthy, competitive relationship among schools - especially in large urban centers. Misinterpreted results can be harmful to everyone involved.

Comment suggested that this is particularly true for 30/33 results when students move in and out of programs, and the overall averages tell little about the success of the school's program. One study participant stated that, "the hardest working teachers I know work in economically, socially challenged schools and have difficulty meeting provincial averages." Interviewees proposed that elaboration and contextualization were required in order for results to be interpreted accurately. FOIPP was identified as a major issue in small schools where often there was only one teacher per subject so that students' results were easily associated with a particular "teacher's performance".

A number of questionnaire respondents indicated that in 1984 when the diploma examinations were reinstated, teachers were assured that the results would never be used as a base to compare schools, classes, or teachers. But many believed that the media and Alberta Learning have in fact contributed to these ends. Overall, participants believed that publishing results on the Internet and availability of data to special interest groups who place select diploma averages in newspapers have made educators cognizant of the importance of the student results to the public, to school boards, and to school and district administrators. As a result, many educators stated that they had changed their teaching practices. Interviewees stated that they are "... not sure they are producing more successful students all around, but they are better test takers." Many believed that the diploma results reported are not so much the result of

good education, but rather the results of a combination of factors – many of which are beyond the control of teachers and schools.

Some research participants believed that examinations have had a negative effect on teachers as, "too many variables exist that we have no control over." A weakness of the examination program is that it is still a paper-pencil test and a "one-shot" for students in terms of how they do on a particular day. Many expressed a wish for a different mark split than the 50%-50%, teacher awarded mark-diploma examination mark, allocation. A number of participants would like to see a lower percentage attributed to the diploma examinations in the belief that a lower weighting would better recognize teacher autonomy. On the other hand, some participants have suggested that in cases where students are not in teacher/school "friendly'" situations, students can significantly pull up their overall subject mark, given the equally-weighted examination mark.

Many respondents believed that good results are related to good teaching. In small schools, often the only scenario was one teacher per subject; therefore, teaching assignments were non-negotiable. Respondents from large schools suggested that some of their colleagues have elected not to teach diploma courses because of the external examination aspect. Some teachers believed, "it is unfair to evaluate a year's or semester's work in 3.5 hours." One former 30 level teacher, who is now an administrator, said that, "A lot of our stronger teachers do not want to teach diploma subjects. I have always thought it was a perk to be a diploma teacher, but some teachers do not think it is a perk at all." Many of those interviewed believed that the obvious reward for a job well done was to give teachers more diploma classes. Many research participants suggested that, "there is no additional support for diploma teachers– in fact the opposite is true. Diploma classes can be larger than option classes and have fewer computers and resources available." Many of the teachers

interviewed, stated that they were not privy to evaluation information or how administrators use or do not use diploma results and that the use of results was an ongoing concern.

Administrators believed that they are also being evaluated when preparing annual reports for the school, district, and the public. Although such evaluation was informal, they recognized it added stress to their job. Principals had been praised for good results and asked to explain poor ones. An administrator commented that, "I would not place an interim or new teacher in a diploma course . . . I would not risk it." This was an indication of the real concern over results expressed by administrators. Teachers also supported this perception, stating that it was unlikely that a temporary teacher would be asked to teach a diploma course given the significance of the marks. In the open-ended comments, participants cited situations in which teachers with poor examination results had been removed from teaching diploma courses or had been encouraged into early retirement. Particular cases of interim or temporary teacher contracts being dependent on good diploma results were also identified.

Many respondents suspected that administrators looked at student performance and determined teaching assignments based on anticipated diploma marks. Diploma teachers noted that only the 30/33 teachers were commended for "a job well done" when results were released–not the prerequisite course teachers. A number of respondents believed that teachers were judged, to some extent, by how well their students did on diplomas, ". . . who gets the best classrooms, timetable, the option of picking their teaching assignment, and students, appeared dependent on the quality of the student marks achieved. Good diploma teachers are valued and respected in the school." Poor marks, by comparison, resulted in teachers being given fewer, if any, 30-level teaching assignments. This was especially true in large schools

and in private schools where staff could be easily moved. An interviewee stated that, "Staff have been declared surplus in our district because of poor diploma results and there has been more than one instance. Entire departments have been dismantled because of persisting diploma problems."

Professional development activities are believed to be loosely linked to improving diploma marks. Unless there is an evident problem, research participants believed that schools used a global approach to improving student learning rather than diploma-specific endeavors. Teachers decided whether professional growth plans, which have replaced formative evaluations, included student achievement goals for diploma examinations. A number of respondents from large schools cited a past experience with their board, whereby teachers were instructed not only to include diploma goals, but were told to set a particular percentage targets in their growth plans. These teachers also believed that were not given any direction or assistance in achieving this district goal. A few study participants suggested that the reason for little administrative intervention was that some administrators view mentorship as an infringement on teachers' professional autonomy. Others expressed leadership concern in that, "our superintendent has said in public that achievement is a priority in our school district. This has not been translated into appropriate practice–yet results are still being over analyzed."

Most interview participants believed that teacher supervision included an aspect of student assessment, albeit an informal one. According to study respondents, verbal recognition by administrators was common, but no record of student examination achievement was placed in teachers' permanent files. Student results have been useful to teachers who, as professionals, used them to gauge their students' progress and the success of their teaching strategies. Many research participants believed that ultimately classroom instruction had to

improve if teachers want students to perform better on the diploma examinations.

All respondents agreed that the level of stress associated with teaching a diploma subject has increased over the past five years. The cited sources of the stress included: the increased accountability of diploma teachers that is "very public and in your face;" the publishing of "out of context" results by the media; the ranking of schools; teaching diploma courses without understanding the curriculum standards; pressure to get students into particular post secondary programs; a high level of responsibility and accountability without a lot of control over the "product" or the "raw materials"; goals that teachers had self-inflicted on themselves to improve on past performances; concern over covering the curriculum adequately; elimination of optional course content due to time constraints; not knowing how, or if, the diploma results were used for evaluation; and being forced to teach diploma subjects because they are the only one with the "expertise." One respondent expressed frustration stating, " I have offered to help students several times and only two out of 70 students took me up on my offer. At the same time, I am held accountable for their results."

Non-diploma respondents believed that they had the professional autonomy to assign final course marks. "When I give a final mark, it is the final mark. No one is going to second guess me. I have failed up to 30% of my students for underachieving and the administrators do not know my program well enough to offer constructive criticism. The buck stops here!" Participants believed that non-diploma subject teachers have more autonomy in what and how they teach. They believed that a number of administrators lack the expertise to adequately scrutinize all the courses in their schools.

Reducing teacher workloads, not necessarily reductions in class size, would help reduce the stress felt by diploma subject teachers. Teachers cited

the discrepancy in teaching load in the province as problematic. They recognized that most teachers in the province are assigned 7 out of 8 classes while a few taught only 6 out of 8. This was identified as a problem considering teachers performed ultimately the same job and received approximately the same pay. A reference was made to future teacher shortages in the mathematics and science areas with the suggestion that the "changing of the guard" would have serious student result implications. Many respondents were concerned that high school has become more stressful in general and that the added stress associated with teaching diploma subjects has contributed to teacher burnout.

Teacher mark advantage/disadvantage was an issue identified in some schools. The discrepancy between teacher and diploma awarded marks was problematic: "A former principal criticized teachers if marks, were a) higher than the diploma mark, b) lower than the diploma, or c) the same – teachers cannot win!" A number of study participants recognize the correlation of the two marks as " a fluke, at best!" Although a few participants described the standards as "arbitrary yet desirable," they raised a concern that Alberta students have to achieve a higher standard than students from other provinces due to the rigor of the diploma examinations; therefore, they were at a disadvantage in competing for quota faculties in Alberta and abroad. Questionnaire comments indicated that a 70% average in Alberta was not the same as a 70% from other provinces and that this difference should be considered in post-secondary admissions.

## STUDY COMPARISON

This final section of the findings reviews the highlights of the 1993 Loerke study followed by a presentation of the comparative findings in this study on a item by item basis. The findings address the last research question: To what

extent have teachers' perceptions of the role of diploma examinations changed since the Loerke(1993) project?

1993 Research Results

Seven years have elapsed since the 1993 Loerke study and this research study was conducted. The findings of the Loerke (1993) study are of interest to teachers, especially those directly involved with diploma courses. A significant number (30%) of the administrators and teachers surveyed believed that an informal use of student results in teacher evaluation was prevalent. The percentage of administrators (1.4%) using student results in formal teacher assessment was somewhat lower. Loerke (1993) suggests that there is a need to be vigilant about this practice over time because of the potential for misuse. How are the student results being used? is one of the questions these findings raise. Another question that emerged from the study is: Why there are marked differences between administrators' and teachers' perceptions in large schools compared to their counterparts in smaller schools in Alberta?

The Loerke (1993) study is valuable as an indicator of the perception that student results were being used in the evaluation of teachers. Of note was the prevalence of informal teacher evaluation based on student diploma examination results.

One of the largest, single, common factors identified in the Loerke (1993) study was the unawareness of teaching staff of the published document, Guidelines for Interpreting and Using the Results of the Diploma Examinations (1993). An updated version of this Alberta Education (1998a) paper contains a brief description of the purpose for provincial assessment along with considerations for interpreting diploma results, factors that affect student achievement, a systematic approach for the effective use of examination results,

and a list of related references. Awareness of this document is fundamental for teachers as it encourages reflective teaching activities and reduces the potential for inappropriate assessment practices. The more informed teachers are of sanctioned and appropriate uses of student results, as advocated by Alberta Learning, the less likely the potential for misuse.

Loerke (1993) predicted that school administrators held significantly different views regarding the use of student results in teacher evaluation than classroom teachers. He also hypothesized that school size should have no effect on participants' perceptions. Loerke (1993) found, however, that the larger the school, the greater the uncertainty of the use of student results. He identified statistically significant differences between teachers' and administrators' perceptions regarding transfers, staffing, and job stress. Administrators believed that student results did not affect teacher transfers nor changes in teaching assignment; however, in his analysis a positive relationship between good student results and teacher promotions was found. The questionnaire design was successful in that it established statistically significant correlation among subgroups, thus supporting explanation building and pattern matching.

School size and teachers' role or position were variables that were statistically significant in the Loerke (1993) study of teacher perceptions and they were also significant in this study. Both studies were limited to teachers' perceptions rather than a compilation of case study data. Although perceptions are difficult to substantiate, common beliefs held by respondents are likely to be a reflection, at least in part, of the ways that diploma results are used.

Loerke (1993) Study Comparison

Of the 37 research questions in this study, 14 items had been replicated from the 1993 Loerke study for comparison purposes. Appendix 2 reports the percentage distributions of research participant responses on the 1-5 perception scale for each of the common questionnaire items. The following comparisons have been made in general terms, i.e., exploring the change in percentage distribution between the studies and identifying any apparent trends in the data. Given the way in which Loerke(1993) grouped his data for analysis, and the factor solution method used in this study, a comparison between common questionnaire items has been based on the distribution of responses on the 1-5 scale for each item only, rather than comparisons based on the reporting categories derived from factor analysis.

To highlight the larger percentage differences between the two studies, the responses on the 1-5 perception scale have been grouped together. The three groups used for comparison purposes were: knew or believed to be true, did not know/unsure, and knew or believed to be false (Figures 10-24). Given the nature of the two studies, items with differences of 5% or more in terms of the percentage of respondents on any of the 1-5 perception scale levels, were considered numerically significant. Differences that were less than 5% were considered insignificant .

Minimal differences were found in each of the 1-5 perception levels for variables 53, 60, 67, and 69. When the perception levels were grouped, the respondents' familiarity with the Alberta Learning published guidelines for interpreting examination results (variable 53), remained numerically greater than the percentage of respondents who were unsure or who had not read the guidelines. Although the change in percentage distribution on each of the 1-5 perception scale choices is not appreciable, the item mean has decreased from

2.35 to 2.16 (Appendix 2). Figure 10 shows a large increase from 60.0% to 69.1% in respondents' awareness of the published guidelines as compared to the Loerke (1993) study.

**Figure 10: Comparison for Familiarity with Published Guidelines (V53)**

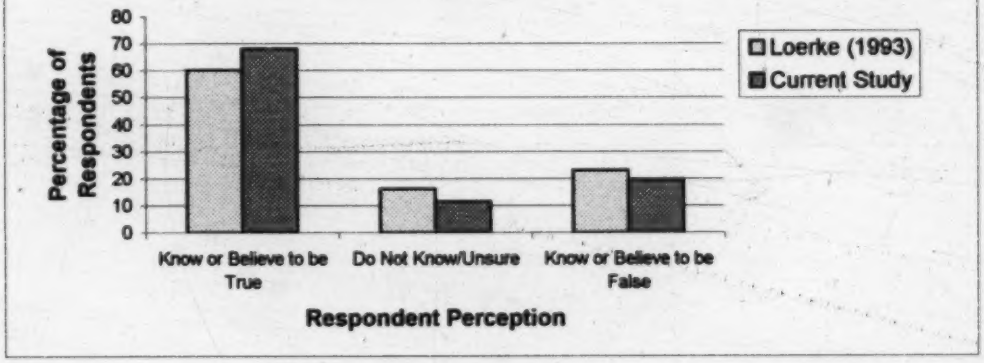



Figure 11 illustrates that minimal changes had occurred in the percentage distribution of respondents for the fair use of diploma results (variable 60). The majority (76.0% in the Loerke study and 78.2% in this study) knew or believed that diploma examination results were used fairly at their schools. Only a 0.7% difference in mean for this questionnaire item appeared between the two studies.

The perception of respondents on the formal use of diploma examination results by the school's administration as part of teacher evaluation (variable 67) remained low. Figure 12 shows that only 7.0% in the Loerke study and 8.3% in this study, knew or believed that this practice was occurring in their schools. Although this percentage has increased slightly, the change in mean for this variable has only been from 3.71 to 3.68, suggesting that the status quo likely exists.

**Figure 11: Comparison for the Fair Use of Examination Results (V60)**

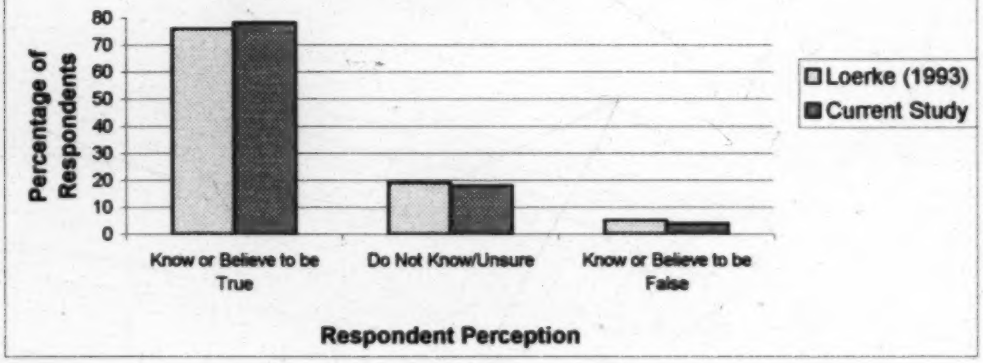
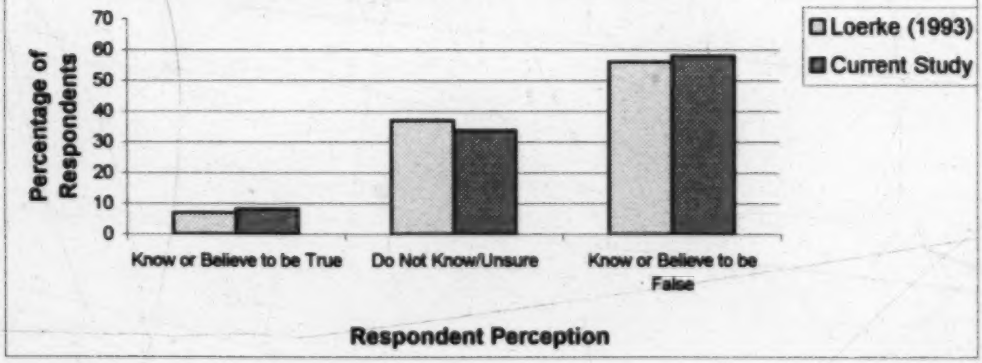
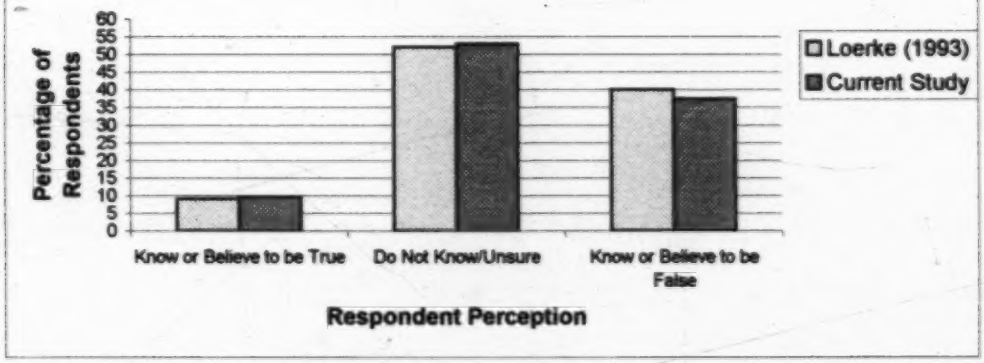



**Figure 12: Comparison for the Formal Use of Results by School Administrators in Evaluation (V67)**



The distribution of responses for the formal use of diploma results by district administrators as part of teacher evaluation (variable 69) were all within 2% of each other. Figure 13 illustrates that only 9.7% of the respondents in this study and 9.0% in the Loerke study, knew or believed that their district administration used student results for this purpose. The difference for this variable is less than 1%, suggesting that no significant change has been reported (not greater than 5%).

**Figure 13: Comparison for the Formal Use of Results by District Administrators in Evaluation (V69)**




Large percentage differences (greater than 5%) on the 1-5 perception scale were reported for variables 54, 55, 66, 68, 72, 74, 75, 78, 76, and 89. Respondents' perceptions on variables 54 and 55 are shown on Figures 14 and 15, respectively. A slightly higher percentage of respondents (from 84.0% to 88.4%) knew or believed that their school's administrators were aware of the published guidelines and resources available for interpreting diploma examination results (variable 54). An even higher percentage difference (from 68.0% to 86.6%) was reported for the district administrator's awareness of the

guidelines and resources (variable 55). The decrease in questionnaire item mean from 1.72 to 1.56 for variable 54, and from 1.96 to 1.69 for variable 55, represents a shift in perception that may reflect the longevity of the current Alberta Learning diploma program. Over time, administrators at both the school and district levels appeared to be better informed now as compared to respondents in the Loerke (1993) study.

**Figure 14: Comparison for School Administrators' Familiarity with Diploma Information (V54)**

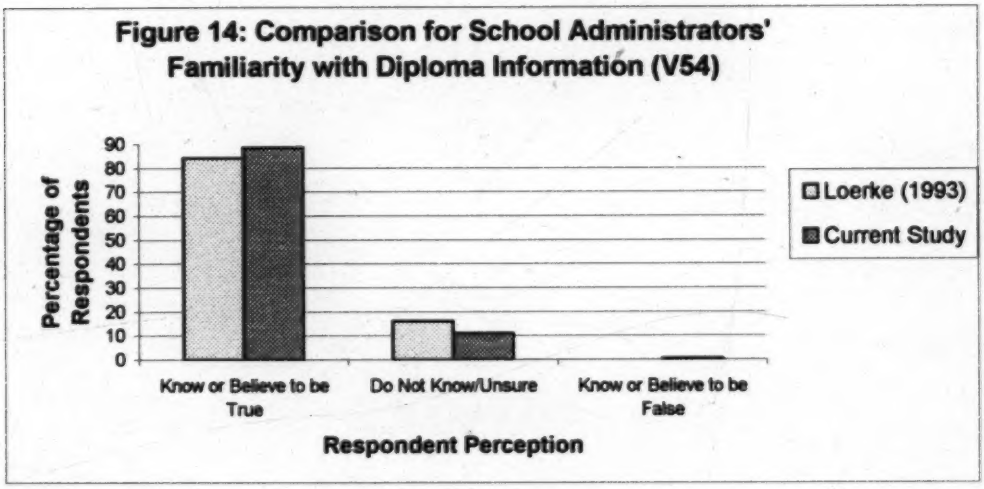
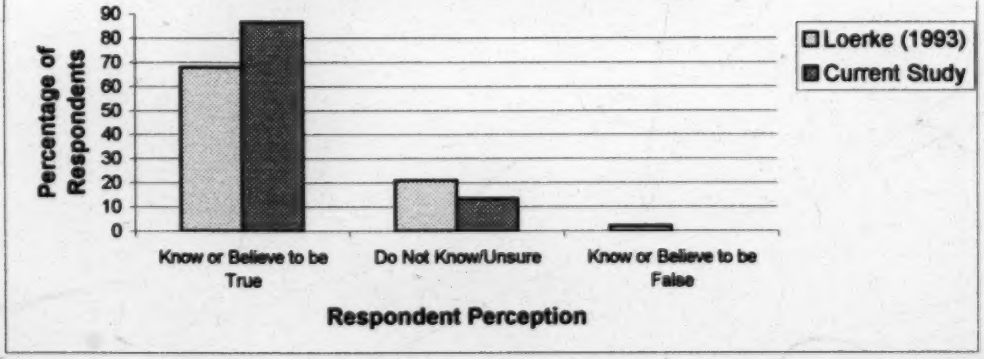



**Figure 15: Comparison for School Administrators' Familiarity with Diploma Information (V55)**



Respondents' perceptions of the informal use of diploma examination results by school (variable 66) and district administrators (variable 68) as part of teacher evaluation are shown on Figures 16 and 17, respectively. In this study there is an appreciable decrease appeared in the percentage (from 33% to 22.6%) of respondents who believed or knew that student results were informally used in teacher evaluation by school-based administrators. This trend is supported by the increase in the percentage (from 30.0% to 35.6%) of respondents who did not believe that diploma results were being used for this purpose. The overall mean for this variable has changed from 2.62 to 3.20. This change in mean is attributable, in part, to the increase in the percentage of respondents (from 37% to 41.8%) who did not know or were unsure as to whether diploma examination results were used informally in teacher evaluation. A similar trend is evident for the districts' informal use of diploma examination results as part of teacher evaluation (variable 66). Although the overall mean has changed only slightly (from 2.99 to 3.08), a significant decrease was observed in the percentage of respondents (from 33.0% to 22.6%) who knew or believed that this practice was occurring in their district. The percentage of uncertain respondents rose by 4.8%.

**Figure 16: Comparison for the Informal Use of Results by School Administrators in Evaluation (V66)**

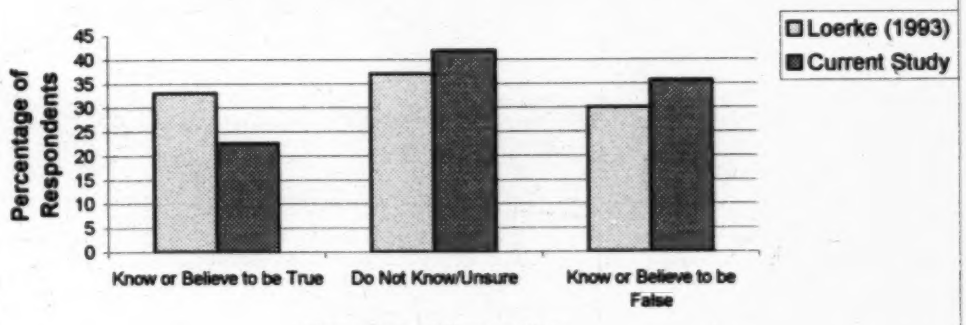



**Figure 17: Comparison for the Informal Use of Results by District Administrators in Evaluation (V68)**

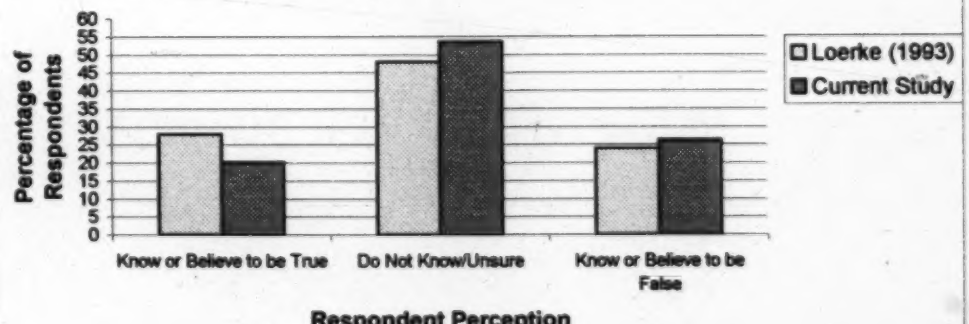
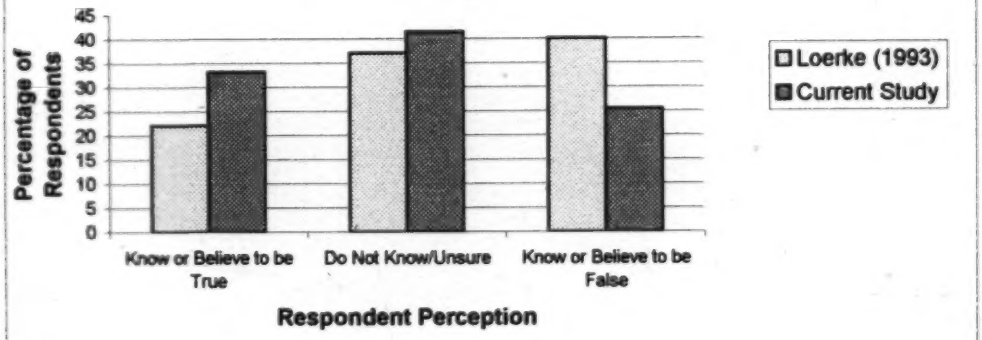


Figure 18 reveals a large increase in the percentage of respondents (from 22.0% to 33.2%) who knew or believed that their school's administrators used diploma examination results to determine teaching assignments (variable 72), as well, a small increase in the number of unsure respondents increased from 37.0% to 41.4%. The mean for this variable decreased from 3.24 to 2.90. Only 25.4% of the respondents in this study, compared to 40.0% in the Loerke (1993) study, knew or believed that diploma results were not used in determining teaching assignments. It appears that more respondents than in the Loerke(1993) study believed that diploma examination results had a direct relationship to teaching assignments.

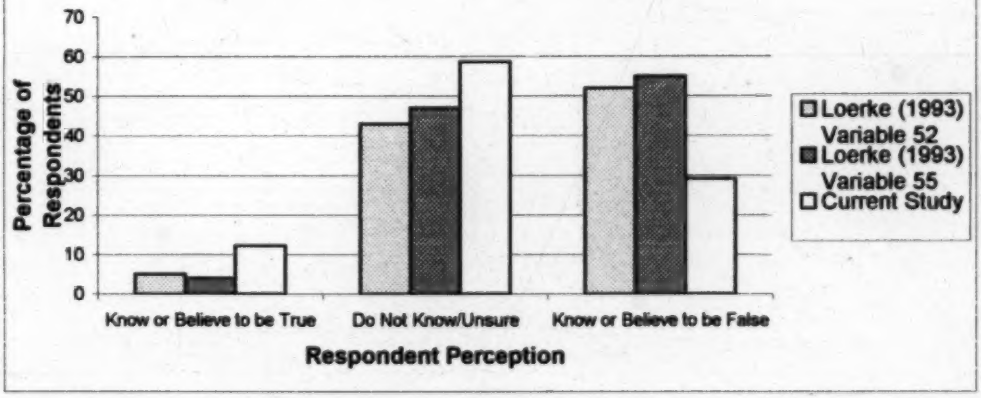
**Figure 18: Comparison for Teaching Assignments (V72)**




Variable 74 corresponds to two variables on the Loerke (1993) study. The results for these two questionnaire items (variables 52 and 55) were very similar as shown in Figure 19. This figure illustrates that the perceptions of respondents in this study who knew or believed that teachers had been

transferred because of poor diploma results had more than doubled (from 4.5% to 12.2%) in percentage This increase is accompanied by a significant decrease in percentage of respondents (from 51.0% to 29.1%) who knew or believed that results were not a consideration in teacher transfers. The percentage of unsure respondents increased (from 45.0% to 58.7%). These distribution changes resulted in a decrease in variable mean from 3.65 to 3.19. Overall, it appeared that the percentage of respondents has shifted from the knew/believe-to-be-false perception to the unsure or knew/believe-to-be-true perception. The majority of respondents in this study did not know if diploma results were a factor in determining teacher transfers.

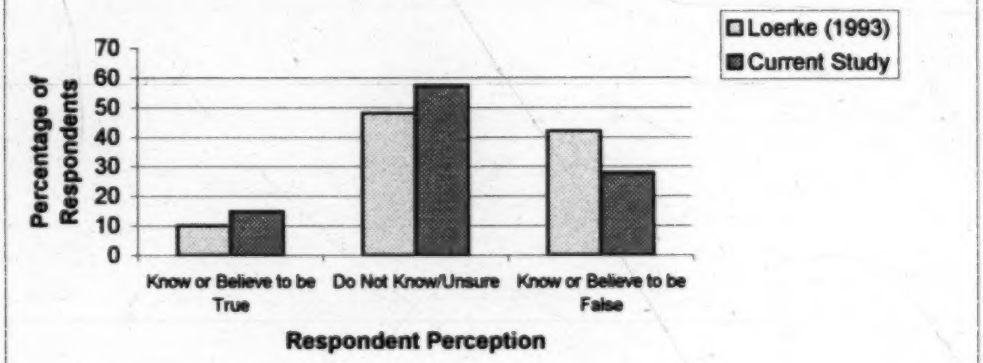
**Figure 19: Comparison for Transfer Requests (V74)**




According to the data in Figure 20, the perception of the practice of promoting teachers to leadership roles because of good diploma results (variable 75) appears to be on the rise. The percentage of respondents (from 10.0% to 14.9%) who knew or believed that this practice was occurring has increased, and a decrease is evident (from 42% to 27.8%) in the percentage of

respondents who knew or believed that this was not the practice at their schools. The percentage of uncertain respondents has increased (from 48.0% to 57.3%) resulting in a net decrease in variable mean from 3.43 to 3.17. Overall, the majority of respondents were uncertain as to whether or not diploma examination results were used as a criteria for school leadership promotions.

**Figure 20: Comparison for Leadership Promotion (V75)**

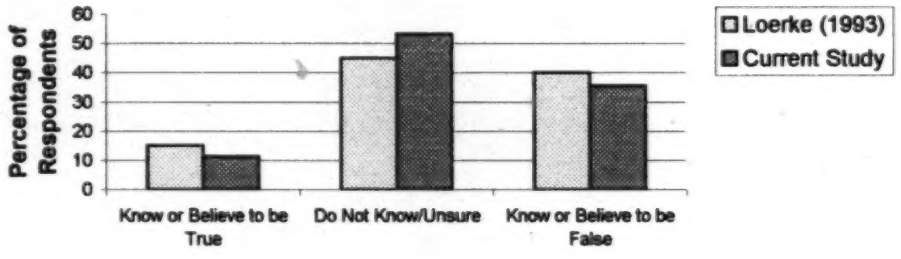



For the two samples, the perceptions of teachers' requests for non-diploma teaching assignments because of poor results (variable 78) has also changed only slightly in most of the 1-5 reporting categories. The variable mean changed very little (from 3.33 to 3.29). Figure 21 shows that the largest difference in perception was reported in the unsure response category (45.0% to 53.3%).

Teacher transfers (variable 74), teaching assignments (variable 72), and promotion of teachers to leadership roles (variable 75), each show an increase in the percentage of unsure respondents. Research participants appear to be

less certain than those in the 1993 Loerke study with regard to the role of diploma examination results and these variables.

**Figure 21: Comparison for Diploma Teaching Assignments (V78)**




Respondents' perceptions of the level of stress associated with teaching a diploma subject compared to teaching a non-diploma subject (variable 76) has increased by 15.1%. Figure 22 shows that 86.2% of the respondents in this study, compared to 71.0% in the Loerke (1993) study, knew or believed that a greater level of stress exists for diploma subject teachers. This change in distribution has resulted in a corresponding decease in variable mean (from 2.20 to 1.72).

Based on data in Figure 23, it would appear that students' ability to select subjects based on their anticipated success on diploma examinations (variable 89) is on the rise. A substantial increase in the percentage of respondents(from 14.0% to 55.4%) who knew or believed that students made course choices based on anticipated diploma results is shown in Figure 23. The percentage of unsure respondents reported dramatically decreased (from 40.0% to 25.9%), resulting in a numerically large change in variable mean (from 3.47 to 2.56).

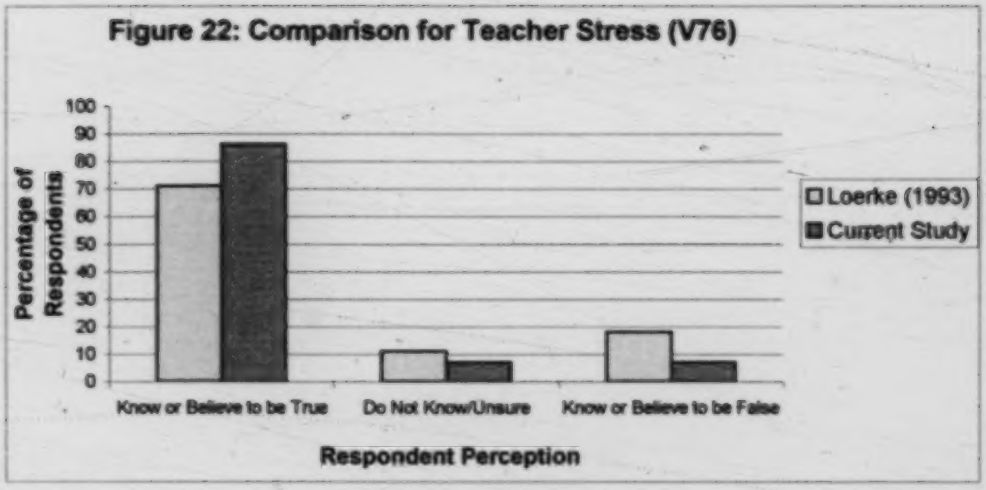


Figure 22: Comparison for Teacher Stress (V76)
Percentage of Respondents
100
90
80
70
60
50
40
30
20
10
0
Know or Believe to be True
Do Not Know/Unsure
Know or Believe to be False
Respondent Perception
Loerke (1993)
Current Study

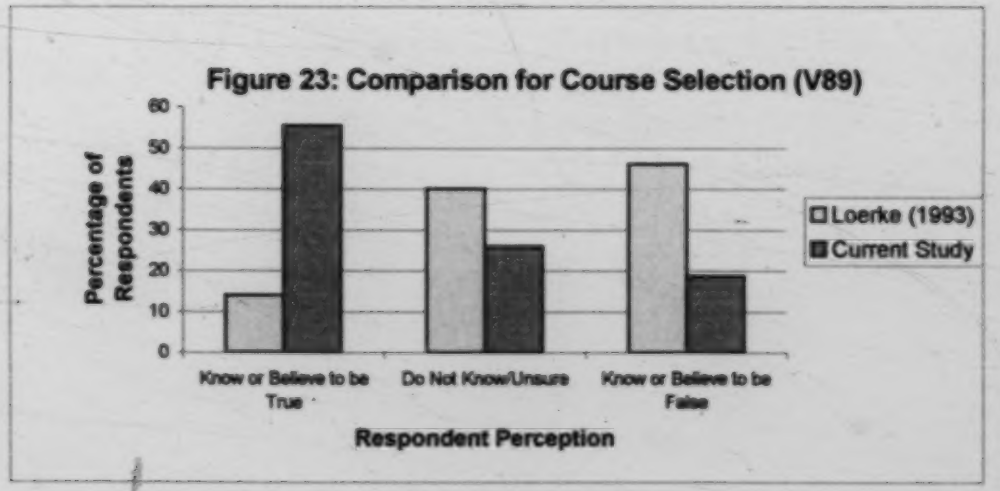


Figure 23: Comparison for Course Selection (V89)
Percentage of Respondents
60
50
40
30
20
10
0
Know or Believe to be True
Do Not Know/Unsure
Know or Believe to be False
Respondent Perception
Loerke (1993)
Current Study

Summary of the Study Comparison

Of the fourteen questionnaire items common to both studies, there are ten items for which there are numerically large differences (greater than 5%) in participants' responses. A much higher percentage of school and district administrators in this study's sample were aware of Alberta Learning guidelines and resources available for interpreting diploma examination results (variables 54 and 55) than in the Loerke (1993) study. More of these respondents were also unsure of the ways in which student results are being used informally by school and district administrators as part of teacher evaluation (variables 66 and 68). Although fewer knew or believed that this was a teacher evaluation practice in their school or district, it appears that more respondents believed that diploma examination results had a direct relationship to teaching assignment (variable 72). A similar trend was observed with respect to the role of diploma examination results in the practice of promoting teachers to leadership roles (variable 75) and in teacher transfers (variable 74). An increase was found in the percentage of respondents who were unsure or believed that diploma examination results played a role in teacher transfers. The substantial increase in percentage of respondents who knew or believed that students were making conscious course choices based on anticipated examination success (variable 89), as well as the other variables previously discussed, culminated in the results reported for variable 76. The large increase found in teacher stress was consistent with many of the changes in perception reported by interviewees and in respondents' questionnaire comments. Many respondents in this study appeared less certain when responding to each of the same items found in the 1993 Loerke study. This increased level of uncertainty is explored further in Chapter 5.

## SUMMARY

The factor analysis provided an overview of respondents' perceptions of the variables studied to address the first research question: How are diploma examinations results used by Alberta teachers? Item means were used to identify potential trends in groups of variables in order to address the ways in which diploma examination results were used. Most of the factor means and individual item means were close to 3 which indicated that respondents did not know or were basically uncertain regarding the questionnaire items. How participants responded on the 1-5 perception scale for individual questionnaire items provided a better insight into the attitudes and beliefs of high school educators. Overall, Factor 4–Assessment Expertise had the lowest item means, indicating that survey respondents knew or believed that educators were aware of the published guidelines and resources available for interpreting results. Factor 1–Staff Evaluation and Factor 8–Improving Results had factor means closest to 3 on the perception scale suggesting that variables in these factors were areas of high uncertainty for research participants. The qualitative findings related to the first research question suggested that strong support for the diploma examination in Alberta exists. Overall, respondents believed that the misuse of diploma examination results had decreased and that the results were used fairly at their schools. Teachers recognized some of the limitations of a single assessment tool that is used on a one-time basis. Concerns were raised that the examinations are used to measure more than they were designed to.

The quantitative analysis examined the perceptions of several educational subgroups including teacher category, school position, school size, and school role. This data was used to answer the second research question: What is the relationship between the perceptions of various educational groups regarding the role of diploma examinations? Using *t* tests and ANOVA resulted

in the identification of statistically significant differences were found among subgroups. In eight of the nine factors studied, statistically significant (at 0.05 or 0.01 level) within subgroups were identified. No statistically significant differences were found for Factor 3–Course Selection.

In Factor 1-Staff Evaluation, administrators reported the most polar views, believing that examination results were less likely to be used in staff evaluation than did diploma or non-diploma teachers. More diploma subject teachers believed that the results had greater impact on instruction than did their non-diploma counterparts. For Factor 1–Staff Evaluation and Factor 2–Impact on Instruction, school position and school size appeared to have little or no effect on respondents' perceptions. Factor 4–Assessment Expertise, Factor 5–Teacher Stress, and Factor 7–Results Recognition generated statistically significant differences for each of the selected reporting categories. These results indicate that large differences in respondents' perceptions exist. In Factor 6–Staffing Consequences, administrators and respondents in small-sized schools were more likely to believe that transfers and teaching assignments were not related to poor student examination results. Only diploma and non-diploma subject teachers reported statistically significant differences for Factor 8–Improving Results. Diploma teachers were more certain regarding the strategies used to improve results than were their non-diploma counterparts. School size was the only subgroup identified as statistically different for Factor 9–Use of Results. Respondents from large schools reported lower means, thus more certainty in the perceived use of results identified in this factor than did those from medium-sized schools.

Qualitative data was used to answer the third research question: How well informed are high school educators regarding the legitimate use of student test results? A number of variables related to the familiarity of Alberta teachers

with the legitimate use of diploma examination results were identified in the content analysis of the qualitative data. Interview data and questionnaire comments provided a more detailed description of teachers' perceptions of the appropriate use of student test results than was possible with the questionnaire alone. The headings selected to address this research question included: awareness of Alberta Learning documents, assessment training, results sharing, school/district policy, school selection, the effect of the media, understanding demographics, and student/school ranking.

Overall, research participants were familiar with most of the Alberta Learning documents and guidelines, but lacked the formal training necessary to make the most of the available information and resources. School/district policy on course admission varied across the province. Although concerns were expresses regarding the negative effects of the media, a number of respondents indicated that there was also positive effects. Research participants' understood that demographic factors affected student performance, many of which were beyond the control of the classroom teacher or school-based administrator. The ranking of students/schools without consideration of the demographic factors was viewed by many research participants as problematic. The criteria for school selection was partially dependent on school/program availability, but was often based on where a student's friends attended school.

There were a number of issues raised surrounding the use fourth research question: To what extent are student test results being used as an endorsement of teacher competency? The qualitative data indicated that teachers/administrators were being held accountable for results without regard for the many variables over which they had no control. Study participants believed that accountability for student examination marks was, therefore, somewhat unjustifiable. A number of the respondents believed that teachers

made a difference in the students' ability to perform well on examinations and that a relationship between "good teaching" and "good results" existed. The "teaching-results" relationship was believed to be more meaningful if studied over a number of school years, rather than for a single examination administration. Research participants indicated that professional development, teaching assignment, assessment expertise, and teacher stress need to be considered if student test results were to be used as an endorsement of teacher competency.

The extent that teachers' perceptions of the role of diploma examinations have changed since the Loerke (1993) project, was the subject of the last research question. The study comparison resulted in the identification of a number of numerically significant differences in participants' responses to questionnaire items. Ten of the 14 questionnaire items common to both studies showed large differences (greater than 5%) in respondents' perceptions. In general, research participants in this study were less certain compared with those in Loerke's 1993 study. The largest perception differences reported were in teacher stress (variable 76) and in student selection of courses based on anticipated examination success (variable 89). Overall, a change in Alberta high school educators' perceptions on the role of diploma examination results was evident.

# CHAPTER 5

# SUMMARY, CONCLUSIONS, AND IMPLICATIONS

## OVERVIEW

This final chapter presents a summary of this study's purpose, a review and discussion of the findings, conclusions drawn from the findings, implications for research and practice, and a final comment.

## PURPOSE OF THE STUDY

The purpose of the study was to identify the perceptions of Alberta teachers on the role of diploma examination results. This general purpose gave rise to five research questions:

1. How are diploma examination results being used by Alberta teachers?
2. What is the relationship between the perceptions of various educational groups, such as: school-based administrators versus teachers; diploma subject teachers versus non-diploma subject teachers; and among teachers in small, medium, and large-sized high schools, regarding the role of diploma examinations?
3. How well informed are high school educators regarding the legitimate use of student test results?
4. To what extent are student test results being used as an endorsement of teacher competency?
5. To what extent have teachers' perceptions of the role of diploma examination results changed since the Loerke (1993) project?

## SUMMARY OF THE FINDINGS

In this section, the findings of the study are restated for each of the research questions: (1) the perceived role of diploma examination results, (2) the perception differences among respondent categories, (3) the assessment expertise of Alberta educators, (4) the use of student results in teacher assessment, and (5) the Loerke (1993) study comparison.

### 1. Perceived Role of Diploma Examination Results

There were two sources of data for used for addressing the first research question. Quantitative data was obtained from the factor analysis of the questionnaire items. Qualitative data was compiled from the ten interviews conducted and the open-ended comments teachers made on the questionnaire.

The results of the factor analysis provided insights into the ways diploma examination results are perceived to be used in Alberta high schools. Nine factors emerged form the factor analysis. The factors are: staff evaluation, impact on instruction, course selection, assessment expertise, teacher stress, staffing consequences, results recognition, improving results, and the use of results. Of the nine items in Factor 1–Staff Evaluation, most respondents were basically uncertain if student results were used as a part of formal or informal staff evaluation or if these practices had increased in the past five years. They did however, report more certainty that student results were not used formally in staff evaluation. Although respondents reported greater certainty in Factor 2–Impact on Instruction, an overall factor mean of 2.72 suggested that basically they were uncertain about the role of diploma examinations in classroom instruction. The item 15 mean indicates that respondents knew or believed that the diploma examination program had been instrumental in aligning curriculum, but had also reduced teacher autonomy. A similar mean for Factor 3–Course

Selection of 2.63, indicated that respondents did not know or were unfamiliar with course selection trends. In Factor 4–Assessment Expertise, individual item means show that the majority of respondents knew or believed that they were well informed regarding guidelines, resources available, various government documents, and examination reports used to interpret results. Research participants were unsure if they had received adequate professional training in assessment of student achievement. Of those who responded to this Factor 4 item, 40.3% knew or believed they had received adequate training and 40.3% did not. Results for Factor 5–Teachers Stress, showed that respondents knew or believed that greater stress was associated with teaching a diploma subject than a non-diploma subject and that no rewards or perks were given for "good" diploma results. Data for Factor 6–Staffing Consequences, indicated that participants were unaware if teachers' requests for transfers or for non-diploma teaching assignments were a result of poor diploma results. The Factor 6 mean of 2.67 indicated similar uncertainty; however, the mean for item 14 indicated that respondents knew or believed that the administrators shared the details of diploma results with all teaching staff. Item means near 2.5 for Factor 7–Results Recognition, indicted a higher level of certainty than a number of other factors studied. Teachers were recognized for their contribution toward desirable results, and respondents knew or believed that the media had little or no effect on the ways in which results were used. The factor mean of 3.12 for Factor 8–Improving Results, indicted that respondents were basically uncertain if teachers were provided with support resources and mentors, if recognition was given to teachers who aligned their school and diploma marks, or if student results were used in professional growth plans. In Factor 9–Use of Results, diploma marks were believed to be a criterion used by students or their parents to select high schools.

In addition to the factor analysis of questionnaire items, interviews and open-ended questionnaire comments provided relevant information regarding each of the factors. Based on this qualitative data, most respondents did not believe that students' diploma results were used formally in staff evaluation. The informal use of results in staff assessment was reported to be an unknown by many respondents; however, the use of student results as part of teacher assessment is believed to have increased over the past five years.

In terms of Factor 2–Impact on Instruction, interview and qualitative comments indicated that the majority of respondents believed that classroom instruction had improved as a result of the reintroduction of diploma examinations in 1984. The examination program had a positive effect on student performance, was instrumental in aligning high school curriculum in Alberta, and had improved classroom teacher assessment practices. Teachers' professional autonomy appeared to be somewhat intact, although a number of respondents believed it had been reduced by the use of diploma examinations.

Qualitative data for Factor 3–Course Selection, showed that respondents in this study believed students were encouraged to make course selections which were in their best interests. Although many respondents reported grade-point entrance requirements greater than the 50% existed at their schools, most believed that the school policy was used only as a guideline for students and their parents to determine the preparedness for subsequent courses. The 30-33 course enrollment trends were not well known by survey respondents.

For Factor 4–Assessment Expertise, the qualitative and quantitative data indicated that district administrators, school-based administration, and diploma teachers were well informed regarding the Alberta Learning publications for interpreting and analyzing diploma results. Interview and questionnaire respondents were unsure if teachers' and administrators' professional training

was adequate for assessing student achievement and recognized that assessment expertise was required if educators were to make the most of student assessment data.

Teacher Stress, Factor 5, was considered by many interviewees to be a given. They recognized that there was more stress for teachers, students, and administrators associated with diploma subjects than non-diploma subjects. The stress appeared to be associated with the ways results were used which is explored further in the third research question.

For Factor 6–Staffing Consequences, many of the questionnaire respondents did not believe that teaching assignments or teacher transfers were directly related to diploma examination results. Most interviewees, on the other hand, suggested that teaching assignments were not only related to teacher expertise, but that teachers with good diploma results were "'rewarded" by being given even more diploma courses to teach. Research participants believed that the renewal of temporary teaching contracts was dependent on examination results. The majority of respondents knew or believed that administrators shared results with all high school staff by providing course averages only for non-diploma teaching staff and detailed class-by-class data for diploma subject teachers.

In terms of Factor 7–Results Recognition, questionnaire respondents suggested that recognition was given to all teachers for their contribution toward student results. Interview respondents supported this belief, but noted that it was not common practice to include prerequisite course teachers in diploma results discussions. Many schools relied on department heads or administrators to disseminate diploma result information. The non-diploma teachers interviewed indicated that they found out about their school's results in local

newspapers and that they had little or no involvement with diploma courses or results analysis.

Related to Factor 8–Improving Results, the majority of teachers did not believe that support resources and mentors were provided at their schools to improve diploma examination results. They also did not believe that teachers were encouraged to include specific diploma results statements or goals in their individual professional growth plans. Comments on the questionnaire indicated that teachers believed that school and diploma mark alignment was somewhat controversial. Many believed that the two marks should not be the same as they measured different aspects of the Program of Studies. Others believed that there should be some consistency between the two marks and that large discrepancies "disadvantaged" students in terms of scholarships and post-secondary institution admissions.

Student examination results appeared to be an important criterion used by students and parents when selecting high schools, as was the case for this item in Factor 9–Use of Results. Many of those interviewed, as well as individuals who commented on the survey questionnaire, recognized the importance of results as an significant consideration in judging the quality of a school. Students' friends and peers, however, continued to be the single most deciding factor in students' choice of schools.

Overall, diploma examination results are being used in a variety of ways by Alberta teachers. A number of the uses are beyond the scope of what diploma examinations can measure. For many of the nine factors studied, teachers perceptions appeared to be basically uncertain, which supported the need for further analysis of the data.

2. Perception Differences Among Respondent Categories

A number of perception differences about the role of diploma examination results were found between teachers and administrators, between diploma and non-diploma teachers, and among respondents in small, medium, and large-sized schools.

Statistically-significant perception differences were found between teachers and administrators for five of the nine factors studied: staff evaluation, assessment expertise, teacher stress, staffing consequences, and results recognition. Generally, administrators were more likely than teachers to believe that they were better informed regarding Alberta Learning guidelines and the resources available for interpreting results, and believed that they had received adequate assessment training. Most administrators felt more strongly than teachers that student results on diploma examinations were not being used in staff evaluation, in the promotion of teachers to leadership roles, and in the use of test results as a criteria for contract renewal, or for staff placement. They were, however, more cognizant of the high level of stress associated with teaching diploma subjects than indicated by teacher respondents. Administrators attributed minor significance to the role of the media in determining how diploma results were used at their schools.

School size, as defined by the number of students writing diploma examinations per year, was a statistically significant contributor to perception differences among respondents. Of the nine factors studied, five were identified as having statistically significant differences (at 0.01 level): assessment expertise, teacher stress, staffing consequences, results recognition, and the use of results. Respondents from small schools reported a higher level of assessment expertise, teacher stress, and results recognition. Respondents in medium-sized schools reported less teacher stress, but more uncertainty in the

use of results. Teachers from large schools reported lower levels of assessment expertise, teacher stress, and results recognition, compared with their small school counterparts. A higher level was reported by large school staff members over those in medium-sized schools for the use of results factor. Staff consequences, such as teacher transfers and allocation of teaching assignments as a result of poor diploma results, were least likely to be reported in small schools than in large or medium-sized schools.

Statistically-significant differences in perceptions between diploma and non-diploma subject teachers were found in five of the nine factors in this study. Significant differences (at 0.05 level) were identified in assessment expertise, results recognition, and improving results. Significant differences (at 0.01 level) were identified for the impact of results on instruction and for teacher stress. Diploma-subject teachers reported significantly higher levels for the impact of diplomas on classroom instruction, assessment expertise, teacher stress, and results recognition than did non-diploma subject teachers. Generally, non-diploma teachers neither knew the impact of diploma examinations nor the various strategies employed to improve results.

Factor 3–Course Selection is the only factor for which no statistically significant difference was found. Eight of the nine factors studied showed significant differences among respondents' perceptions on the role of diploma examinations.

3. Assessment Expertise of Alberta Educators

Factor 4–Assessment Expertise was one of the nine factors identified by factor analysis. T-tests and ANOVAs were conducted to further analyze the reporting groups identifies by the researcher.

Assessment expertise was studied using three groups of respondents: administrators versus teachers; diploma versus non-diploma subject teachers; and small versus medium versus large-sized schools. A significant difference (at 0.01 the level) was found between administrators' and teachers' perceptions. Administrators reported higher levels of assessment expertise, in that they knew or believed that teachers and school-based/district administrators were aware of published guidelines and resources available for interpreting diploma examination results and that diploma teachers used various reports to analyze results. Administrators felt more strongly than teachers that educators received adequate professional training in the assessment of student achievement.

Diploma subject teachers more often knew or believed that educators had more assessment expertise than their non-diploma teacher counterparts. Statistically significantly differences (at 0.05 level) were found between diploma-subject teachers and non-diploma teachers. Respondents in small schools reported higher levels (significantly different at 0.01 level) of assessment expertise for this factor than respondents in large schools. Based on the means for this factor, administrators were the group who most often believed assessment expertise was adequate. Although most respondents knew or believed that the administrators shared the details of the school's results with all teaching staff (variable 58), uncertainty exists as to whether teachers/administrators received adequate professional training (variable 94). The standard deviation of 1.10 and mean of 3.03 suggested that respondents varied greatly in what they believed was adequate assessment training. Internet access to examination information was believed to have a positive effect on diploma results (variable 57), although comments on the questionnaire suggested some problems were evident in teacher access to assessment information at the school level.

Of the 402 respondents, 86.1% knew or believed that there was a greater level of stress associated with teaching a diploma subject than a non-diploma subject (variable 76). According to interviewees, some of this stress was related to teachers' unfamiliarity with Alberta Learning *Guidelines for Interpreting Results* (variable 53). Over 69% of the questionnaire respondents reported having read these guidelines, but a standard deviation of 1.47 suggests that a number had not.

Assessment expertise appears to be one of the most divergent factors studied based on the perceptions found among the three groups of respondents. Although the majority of research participants appear to be familiar with the published guidelines and resources available for interpreting diploma results, the lack of adequate professional training in learner assessment and the increased stress associated with teaching diploma subjects are of concern.

4. Use of Results in Teacher Assessment

Factor 1–Staff Evaluation was one of the nine factors generated by the factor analysis of the questionnaire items. Nine of the 37 questionnaire items, most of which relate to the formal and informal use of results by district and school administrations, loaded on this factor.

Generally respondents knew or believed that diploma examination results were not being used, either formally or informally, as part of teacher assessment. A statistically significant difference (at 0.01 level) was found between administrators' and teachers' perceptions whereby administrators felt more strongly that results were not being used in teacher evaluation. The highest standard deviation was observed for the formal use of results as part of teacher evaluation (variable 66), suggesting a number of divergent perceptions by respondents. The majority of respondents were basically uncertain if diploma

results as part of teacher assessment had increased in the past five years (variable 82). Means for Factor 1–Staff Evaluation, were at, or near, three, suggesting that the majority of respondents did not know or were unsure of the role of diploma results in staff assessment. Only 8.3% of the respondents knew or believed that student results were used formally by school administrators in teachers evaluation (variable 67) compared with 22.6% who believed results were used informally (variable 66).

Staffing consequences such as teacher requests for transfers and non-diploma subjects because of poor results, were generally not believed to be true. Statistically significant differences (at 0.01 level) between administrators and teachers was found. Administrators appeared more certain than teachers that these practices were not occurring at their schools. Generally, respondents believed that diploma examinations were not used by their school's administration to determine teaching assignments (variable 72). Although teaching assignments were somewhat based on subject expertise, interview respondents suggested that placement was also student results dependent. When asked if diploma results were used fairly at their schools (variable 60), collectively, respondents knew or believed that this was true. Whatever practices or policies are in place, most teachers perceived them to be fair.

The use of diploma examination results in teacher assessment appears to be limited. The informal use of student results was more prevalent than the formal use of results in staff assessment. Administrators were more certain than teachers, and diploma teachers more certain than non-diploma teachers, about the use of examination results in assessment-related practices.

5. Loerke (1993) Study Comparison

The intent of the Loerke (1993) study was to set a benchmark from which further research could be conducted. Although his study established that the use of students' achievement data to assess teachers does exist, it did not appear to be a widespread practice. Informal use of results, by the school-based administrators or by the district, was much more prevalent than the formal use of results in teacher evaluation.

The Loerke (1993) study revealed that teachers and administrators rarely shared the same perceptions. In his study, department heads may have been included under the administrators category and this may have influenced the results. Although this is not the case with the present study, the differences in perceptions appear to be consistent. Of the nine factors used to explore the perceived role of diploma examinations in this study, five factors were identified with statistically significant differences (at 0.01 level) between administrator and teacher respondents. Administrators did not believe that examination results were used to evaluate teachers and principals, to promote teachers to leadership roles, or as a criteria for contract renewal. They also reported significantly higher levels of assessment expertise and recognized the increased teacher stress associated with teaching diploma examination subjects. The majority of respondents in both studies reported that teaching externally-examined courses was far more stressful than teaching non-diploma courses. Teachers tended to be less polar than administrators in their views, with many reporting in the "uncertain" category.

School size played a role in the way diploma examination results are used. In the Loerke (1993) study, teachers in small schools reported greater certainty about the use of student examination results. They were more likely to believe that student results were not being used in an evaluative manner than

teachers from large schools. For Factor 1–Staff Evaluation, no statistically significant differences based on school size were identified in this study. Teachers in small schools were, however, more confident than their medium and large school counterparts that negative staffing consequences, such as transfers, were not occurring at their schools. They also reported higher levels of assessment expertise, results recognition, and teacher stress.

Many respondents in the present study appeared less certain when responding to each of the same items found in the 1993 Loerke study. Of the fourteen questionnaire items common to both studies, ten items showed numerically large differences (greater than 5%) in respondents' perceptions. More respondents were unsure of the ways in which student results were being used informally by school and district administrators as part of teacher evaluation (variables 66 and 68). Although fewer knew or believed that this was the practice in their school or district, it appears that more respondents believed that diploma examination results had a direct relationship to teaching assignment (variable 72). A similar trend was observed for the role of diploma examination results in the practice of promoting teachers to leadership roles (variable 75) and in teacher transfers (variable 74).

A much higher percentage of school and district administrators were aware of Alberta Learning guidelines and resources available for interpreting diploma examination results (variables 54 and 55). An increase was reported in the percentage of respondents who were unsure or believed that diploma examination results played a role in teacher transfers. The substantial increase in percentage of respondents who knew or believed that students were making conscious course choices based on anticipated examination success (variable 89). The large increase found in teacher stress was consistent with many of the

changes in perception reported by interviewees and in respondents' questionnaire comments.

The Loerke (1993) study comparison resulted in the identification of a number of numerically significant differences in participants' responses to questionnaire items. Ten of the 14 questionnaire items common to both studies showed large differences (greater than 5%) in respondents' perceptions. In general, research participants in this study were less certain compared with those in Loerke's 1993 study. The largest perception differences reported were in teacher stress (variable 76) and in student selection of courses based on anticipated examination success (variable 89). Overall, a change in Alberta high school educators' perceptions on the role of diploma examination results was evident.

## DISCUSSION OF THE FINDINGS

In this section, the findings of the study are discussed under the headings of the perceived role of diploma examination results and the Loerke (1993) study comparison.

### Perceived Role of Diploma Examination Results

Compared with the 1993 Loerke study, the increase in the percentage of questionnaire respondents who were well informed of the guidelines, resources available, and government documents and reports used to interpret results, may be a reflection of the general increased emphasis on accountability in education. The substantial increase in the proportion of respondents who perceived that students made conscious course choices based on anticipated examination success also supports this contention. The specific ways in which results have been used in teacher and administrator evaluation appeared to be somewhat

unknown by the classroom teachers surveyed. This may be an indication of the lack of openness of communication among subject area teachers and between teachers and school-based administrators. Although the findings indicate that most administrators shared the results with their high school teaching staff, non-diploma subject teachers indicate that they were unaware of many aspects related directly or indirectly to diploma examination results. These teachers were often not included in the analysis of results and the related goal setting activities generated from the assessment discussions.

Though one important aspect of assessment expertise is the awareness of interpretation documents, knowing how to use the available information is critical if the analysis is to be used to improve student learning. Changes to the teacher supervision policy and the implementation of teacher growth plans may not yet be well understood by the teaching population surveyed. Lacking understanding of the nuances of these initiatives has the potential to contribute to the misuse of student results in teacher evaluation.

Study participants believed that classroom instruction had improved as a result of the introduction and ongoing commitment by Alberta Learning to the diploma examination program, a belief that corroborates the 1990 Calder study commissioned by the ATA. The alignment of curriculum with the Program of Studies and the improvement of teacher assessment expertise may be attributed, in part, to teacher involvement with Alberta Learning in terms of the field test program, diploma marking, and the availability of various diploma-related materials on the Internet.

Concerns about teacher autonomy persist. Interviewees indicated a desire for a balance between professional autonomy and standards, and most approved of the 50-50 weighting of the examination and teacher awarded marks. Although survey participants appeared reasonably well informed regarding the

Alberta Learning publications, interview data and questionnaire comments suggested that the depth of their understanding may be superficial, at best. The focus on means only and the lack of subject matter expertise by school-based administrators, may be indicative of a need for more in-depth understanding, not only in the interpretation of diploma examination results, but in student assessment in general. By all accounts, study respondents lacked the professional training necessary to make the most of the available student assessment information.

The results of poor diploma results appeared to be the reduction of the number of diploma courses assigned to individual teachers, not an investment in teacher professional development with regard to student assessment, curriculum interpretation, teaching resources, or formal mentorship. Teachers with good or superlative student results compared to provincial standards were given more diploma subjects to teach. Study respondents suggested that non-diploma course assignments were more desirable due to the perception that teachers received less scrutiny based on student results. Teachers who could "produce results" got asked and those who couldn't "got off the hook." This "rewarding of incompetence" makes teaching assignments contingent on student results and contributes to a perceived lack of accountability and decrease in teacher responsibility associated with non-diploma subjects. This use of results contributed to the additional stress felt by diploma teachers.

Based on comments and interview data, a number of misconceptions appeared in conjunction with the appropriate use of student examination results. Growth plans should be authored by individual teachers, not by their school district or school administrators (Council on School Administration and the ATA, 1998). School plans, on the other hand, should represent a collaborative effort. Whether or not a teacher-awarded mark and diploma examination mark align

requires contextual information and therefore professional interpretation. School reports are expected not only to provide overall student results, they should outline the contextual factors that attributed to them (*Guide for School Board Planning and Results Reporting*, 2000).

Statistically-significant differences in perceptions were found between administrators and teachers on many of the variables studied. The educators' belief that assessment expertise was adequate was not a view shared by their teaching staff. It is likely that administrators' professional experience and position of authority at the school contributed to their awareness of how student examination results were used in teacher evaluation and in teaching assignment decisions.

School size was also a significant contributor to the perception differences among study respondents. The potential for perceived misuse of results appeared to have a direct relationship with school size. The larger the school, the more likely the perception that teaching assignments were dependent on student results. Teachers in small schools may teach only one subject and their school may be the only choice for local students. This limitation was the largest contributing factor to the perception differences of respondents based on school size. The lower level of teacher stress reported in large schools may be attributed to collegial support, such as: having more than one teacher per subject, making teamwork possible; the anonymity of diploma subject teachers when examination results are reported; and the possibility of staff differentiation whereby teachers specialize in subject matter and course levels. Stress appeared to be school-size dependent. Teachers in small schools, especially those in small communities, were well known by the local residents and therefore felt additional pressure by parents and the community for students to perform at or above provincial standards. By comparison,

teachers in large urban centers could more readily separate school and community life.

The stress associated with teaching diploma subjects was somewhat self evident. Teachers placed heavy expectations on themselves. Students, parents, and the public expected teachers to equip students with the ability to complete high school with the desired courses and at standards that would allow them to compete for scholarships, and to gain post-secondary admission in the institution of their choice. Fear of the "unknown", lack of control over variables affecting examination results, limited assessment and curriculum expertise, increased teacher workloads, and limited communication between administrators and teachers regarding the use of results, especially in staff evaluation, were some of the proposed reasons for the heightened stress levels perceived by diploma teachers and school-based administrators.

Diploma examinations appear to have become the "catch all tools" that the ATA and others warned against since their re-introduction in 1984. Study participants believed that student results were used for purposes other than to credential students and sort them by academic ability. The results were used to publicly rank schools, and to align curriculum across the province; as well they had an impact on teacher selection and teaching assignments. The examinations had been used by Alberta Learning to demonstrate that the curriculum, as far as a paper-pencil test can measure, was taught and that the standards dictated by examination design were met by a minimum of 85% of Alberta high school students.

School authority, school, and individual class result reports detail the expectations and provincial standards used as the benchmark. The ability of stakeholders to interpret the information available and use it to improve student learning for the next cohort of students is questionable. Alberta Learning

suggests that interpretation of diploma examinations be a collaborative effort that involves various stakeholders in assessing the results of the school or instructional group results. Detailed diploma results compare each reporting group to provincial standards on each aspect of the examination. As well, diploma marks and student participation rates for the reporting group compared to the province are provided. Student achievement changes over time and five year participation rates may be more meaningful in measuring student learning than results for a particular school year. It should be noted that diploma examinations are designed to be equivalent in a given school year but not necessarily across school years (Alberta Education, 1998a). This is especially true if a significant modification of the curriculum has occurred. Incentive programs designed to reward high scores that were described in the literature review are somewhat problematic, given that marks from year to year vary slightly without any significant change in the student population. The ways in which diploma results dictate classroom practice requires ongoing study.

Loerke Study Comparison

The results of the present research were comparable in several respects with Loerke's 1993 findings concerning the use of examination results in teacher evaluation. There was little evidence that many teachers perceive that the use of student results in teacher/principal formal evaluation is prevalent. A lack of understanding of the extent of informal evaluation which is based on diploma examination results, however, was apparent.

Administrators and teachers continued to have divergent views on the role of examination results. Diploma teachers in small-sized schools continued to experience higher levels of stress than those in larger-sized schools.

The Loerke (1993) findings and this research study are consistent with

the findings of Nolen, Haladyna, and Haas (1992) who reported a number of significant differences in teachers and administrators' perceptions on the use of standardized test scores. Teachers in the Nolen, Haladyna, and Haas (1992) study were twice as likely as the administrators to believe that results were used to advertise the school and to evaluate teacher and district effectiveness.

Interestingly, in terms of the perceived teacher evaluation practices, no appreciable difference among school-size respondents' perceptions existed. The fairness aspect in the use of examination results remained favorable (up 2.2% in this study from 76.0% in the Loerke study). Although "fairness" was a loaded term, study participants appeared satisfied with the use of student results at the school level. The policies and practices in place were believed to be acceptable by more than three quarters of the professional staff.

Teachers surveyed appeared better informed (up 9.1% from 60.0% in the Loerke study) regarding the published guidelines for interpreting examination results. School and district level administrators were also better informed when compared to the 1993 Loerke study. Perhaps the Alberta Learning Website and the school report requirement of administrators have contributed to the increase in familiarity of government assessment documents.

On several points, however, this study's findings diverge from the Loerke (1993) findings. Appreciable decreases appeared in the percentage of study participants who believed that examination results were used in informal teacher evaluation. However, more teachers simply did not know how the results are used. An 11.2% increase in the percentage of study participants who believed that administrators used diploma examination results to determine teaching assignments and the more than doubled percentage of those who believed teachers had been transferred due to poor diploma results, may be indicative of the push for higher results and the focus on accountability. Administrators

recognized that examination results were affected by teachers' expertise. Selective staffing to enhance student achievement appeared to be a common strategy employed by school-based administrators in the present study.

The stress associated with teaching a diploma course appeared to have increased significantly (from 71.0% to 86.2%). The multiple-uses of student results may be a contributing factor to the stress increase reported by teachers. Teachers believed that students were more savvy in selecting courses in which they anticipated the greatest success. This selection of courses may not, in fact, be at the sole discretion of the student; rather, it may be influenced by classroom teachers and administrators. There is a general perception that the competition for scholarships and post-secondary institutions had raised the bar causing students to be more selective about the courses they take and to repeat courses in order to achieve the desired outcome.

Diploma examinations have had an impact on a number of stakeholders. Heightened accountability is reflected in the results reported by this study's participants. Frank Peters from the University of Alberta recognizes the frustration in the teaching profession and reaffirms what many educators already know:

> The reality of school life today is one of intensification, as more and more tasks and responsibilities are pushed off to the school level, and more and more centrally developed standards and curricula, into which the teachers have minimal input, are imposed. Overworked administrators are challenged to engage overworked teachers in their struggles to respond with sound educational practices in contexts that are becoming more demanding by the day. (ATA News, 2000, p. 6)

The negative effects of increased stress among diploma teachers and the impact on student learning should be explored further.

## CONCLUSIONS

The following generalizations encapsulate the conclusions in this research:

1. Most study participants believed that diploma examination results were being used fairly in their schools.
2. A majority believed that student examinations results were being used to select school staff and to allocate of teaching assignments.
3. A number of the survey participants were not well informed about many aspects related to the role of diploma examination results.
4. Many teachers and school-based administrators believe that they require additional assessment expertise in order to meet the challenge of increased accountability in education.

Within the context of Alberta high schools, the findings of this study provide insight into high school teachers'/administrators' perceptions of the various roles of diploma examinations.

## IMPLICATIONS FOR RESEARCH AND PRACTICE

This section deals with the implications of the study for further research into the topics of external assessment, teacher autonomy, and issues associated with evaluation and implications for practice. In light of this study's findings and the conclusions drawn from them, several orientations for further research and practice concerning the topic of standardized external examinations and teacher autonomy are put forward.

### Implications for Further Research

First, it is recommended that further research be undertaken in order to determine if longitudinal studies of student examination results would be

valuable in studying the effects of various teaching strategies and administrative policies on student learning. Research shows a good deal of stability in a school's effectiveness from year-to-year. The improvement of schools takes place over extended periods of time, consequently, Gray, et al. (1995) suggest the need for studies which track successive cohorts of pupils through their schooling are required if estimates of improvement are to be established. Such research would provide the opportunity to determine the ways in which instructional changes, teacher expertise and experience, and school policy, impact student examination achievement.

Second, in the design of further research, the sample of respondents might be broadened to include high school students. One of the shortfalls of this study, and of a number of earlier studies on diploma examinations, is the exclusion of students. Ultimately, it is their lives that may be affected by the use or misuse of diploma examination results. The number of years they spend completing a high school diploma, the level and variety of academic courses they successfully complete, and the post secondary institutions students qualify to attend, are all diploma examination dependent. It is important to remain vigilant regarding the ways in which high standards may advantage or disadvantage Alberta students seeking post-secondary admission in other provinces or countries. Kohn (1992, 2000) is convinced that external examinations are detrimental to student learning. A study of students' perceptions could be conducted using the same methods employed in this study.

Third, it is recommended that in further research additional forms of measurement standards which might be viable alternatives to diploma examinations, be explored. Such research would be timely, as a performance assessment revolution is underway, according to Asp (1998). "Authentic

assessment" of student learning is but one example of new measurement strategies being discussed. Assessment reform may offer different views on the nature of learning and improvements in assessment technology and methodology. The public has come to expect individual attention and choice in many aspects of life. Why standardized examinations with limited administrations are not designed to accommodate individual students is being questioned. Computer adaptive testing, for instance, is being explored as an alternative to the standardized pencil-paper tests found in diploma examinations. The emergence of interactive environments that facilitate individual growth in addition to serving accountability functions may result in the decline of the conventional, one-time, center-administered examinations. A study of the effectiveness of various student assessment methodologies and technologies compared with the existing diploma examination program may result in better teaching practices and better measurements of student learning.

Fourth, the use of student examination results in school staffing decisions, how teachers improve examination scores, and how teachers and students are encouraged to raise scores, are specific topics for future studies. Case study methodology may be an appropriate means to generate data that can be thematically grouped and analyzed. This research could address a number of specific, unanswered questions arising from this study.

Implications for Practice

The study gives rise to some recommendations for practice. First, diploma subject teachers' stress is an ongoing issue. A number of strategies may be employed to address teacher stress as it relates to diploma examinations. Fear of the "unknown" can be a significant source of stress. Improving the school-level communication between the administrators and all

the teaching staff regarding the use of student examination results in teacher evaluation, teaching assignments, and school marketing appeared to be a province-wide necessity. A mismatch between the perceptions of administration and those of teachers (Loerke, 1993; Nolen, Haladyna, and Haas, 1992) is also evident. Teachers should play an integral role in preparing school reports and setting learning goals if they are to be held accountable for them. The importance of providing a context for school examination results and the education of parents/public in the use of the context in the interpretation of results, are imperative to improving student learning and the health of the teaching staff.

Schools and institutions, in general, should be held accountable for things which they can be expected to influence, rather than for the variables over which they have no control. Teachers, especially those new to diploma courses, may require professional assessment training if they are to use examination results effectively and in order to understand the contribution of those factors over which they have control and of those they do not. Used together, formative teacher evaluation and dependable data on student achievement can represent a strong combination for school improvement, but investing necessary resources is required. Also needed, is a common ground that supports moral and educational practices (Stiggins, 1998; Sirotnik and Kimball, 1999).

FOIPP is an issue for schools and teachers in schools where there is only one teacher for a particular subject and the results published can be directly linked to an individual. Alberta Learning may need to address this as a source of stress, particularly if the Fraser Institute and other forms of media continue to publish school rankings based on specific diploma courses, since such information may have a direct impact on individual teachers.

The integrity of student learning over sequential courses, collegiality, and teacher commitment to student learning may be enhanced by the distribution of teaching assignments among subject area teachers. Teachers who teach the grade ten through grade twelve courses in a subject are better informed about the entire program of studies for that subject area. They are familiar with the content and level of difficulty expected in order for students to be successful from one course to the next. They also have a better appreciation of the many factors which contribute to students' achieving the highest standard possible on diploma examinations. Teachers who are restricted to teaching non-diploma subjects may not have the opportunity or possess the ability to make valuable contributions in the analysis of students' diploma results if they have never taught the course. Teaching all subject levels promotes collegiality whereby all staff are working toward a common goal and they recognize and depend on the feedback from their colleagues. These teachers share the workload, responsibility, and accountability for student learning.

Mehrens (1991) contends that measurement professionals need to clarify what constitutes defensible and indefensible test preparation activities, and then communicate these to teachers. It is possible that if teacher attitudes about the use of diploma examination issues are changed through education, a reduction in the frequency of questionable practices may occur (Peterson and Neill, 1999). District and provincial officials have the right and responsibility to require schools to provide evidence that all students are learning. Teachers and school-based administrators have the responsibility to gain the assessment expertise necessary to improve student learning. Policy makers have a responsibility to support the efforts of educators in achieving this end.

## CONCLUDING COMMENT

In broad terms, this research dealt with the perceived role of diploma examinations in Alberta as perceived by Alberta high school teachers and school-based administrators. This investigation also provided a vehicle of expression through which teachers could express their views, insights, and perceptions related to a standardized external examination. The re-introduction of the diploma examination program in 1984 was an educational reform that addressed accountability and equity issues. Notwithstanding its role as a standards benchmark, it is only a single measure and is therefore limited in scope. Although examination marks are not the "hard evidence" many stakeholders would think and one year's diploma results are not realistic in measuring students' educational growth, the results do have an impact on teachers and administrators.

This study can be considered successful to the extent that it contributes to discussions concerning how diploma examinations are being used and the effect of these uses on student learning. The fact that most of the superintendents and principals in the province agreed to have their high school staff participate in this research study is a tribute to their dedication to understand assessment better and ultimately to improve student learning. Even the principals who did not elect to participate in this study did so out of compassion for their staff. These administrators did not wish to add to the burden of their already overworked staff by asking them to complete another survey, albeit a potentially valuable one.

Performance assessment provides the means for improving student learning, but only if teachers receive sufficient training and support. To this end, AISI is a welcome source of supplementary funding to school authorities throughout the province. One of the goals of this government initiative is to

encourage innovations in practice that have a strong potential, based on research literature, to improve student achievement and to enhance the quality of classroom-based assessment.

We have just begun to pay the dues we owe in the form of the commitment, training, and resources needed to develop the assessment expertise and sound instrumentation required at the school level to assess student learning and evaluate the impact of teachers on that learning (Stiggins, 1989). The diploma program has facilitated this process and, though it is not a perfect tool, has contributed to the establishment of provincial standards and is likely to continue to do so for some time.

# REFERENCES

Alberta Education. (1994). Freedom of information and protection of privacy act. Statues of Alberta. Chapter F-18.5 with amendments in force as of September 1, 1995. Edmonton: Alberta Education.

Alberta Education. (1994). Meeting the challenge: Three-year business plan. Edmonton: Alberta Education.

Alberta Education. (1998). Meeting the challenge IV: The government of Alberta's three-year plan for education, 1998/1999 to 2000/2001. Edmonton: Alberta Education.

Alberta Education. (1998a). Guidelines for interpreting and using the results of the diploma examinations. Alberta education policy 2.26. Edmonton: Alberta Education.

Alberta Education. (1998b). Accountability in education: Use and reporting of results on provincial assessments. Alberta Education policy 2.2.3. Edmonton: Alberta Education.

Alberta Education. (1995). Access to information. Alberta Learning policy 3.2.5. Edmonton: Alberta Learning.

Alberta Education. (1996). School authority accountability. Alberta Learning policy 2.2.1. Edmonton: Alberta Learning.

Alberta Home and School Councils' Association, Alberta School Boards Association, Alberta Teachers' Association, & College of Alberta School Superintendents. (1999). Joint proposal for alternative to school performance incentive program: Alberta School improvement program (ASIP). Proposal to the Alberta Minister of Education. Edmonton, Alberta.

Alberta Initiative for School Improvement Education Partners Steering Committee. (1999). Proposed framework for the Alberta initiative for school improvement. Edmonton, Alberta.

Alberta Learning. (1999). News release: School performance incentive program on hold. Edmonton: Alberta Learning.

Alberta Learning. (2000b). Guide for school board planning and results reporting. Edmonton: Alberta Learning.

Alberta Learning. 2000. Guide for using diploma examination results. Learner Assessment Branch. Edmonton: Alberta Learning.

ATA responds to exam proposal. (1983). ATA News, 17,12, 1.

Asp, E. (1998). Assessment in education: Where have we been? Where are we headed? In Education in a New Era: 1998 ASCD Yearbook. (pp. 123-157). Alexandria, Va.: Association for Supervision and Curriculum Development.

Barlow, M., & Robertson, H. (1994). Classroom warfare: The assault on Canada's schools. Toronto: Key Porter Books Limited.

Berg, B. (1998). Qualitative research methods for the social sciences (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Berk, R. (1988). Fifty reasons why student gain does not mean teacher effectiveness. Journal of Personnel Evaluation in Education, 1, 4, 345-363.

Berliner, D., & Biddle, B. (1995). The manufactured crisis: Myths, fraud, and the attack on America's public schools. Reading, Mass.: Addison-Wesley.

Berliner, D., & Biddle, B. (1999). The awful alliance of the media and public-school critics. The Education Digest, 64, 5, 4-5.

Bishop, J. (1998a). Do curriculum-based external exit exam systems enhance student achievement? Consortium for Policy Research, RR-40. University of Pennsylvania.

Bishop, J. (1998b). High school diploma examinations: Do students learn more? Why? Policy Options, 7-8, 9-13.

Bishop, J. (1999). How provincial diploma exams improve student learning. Fraser Forum, 9, 7-9.

Blackwell, T. (2000, October). Ontario gets a read on student literacy: Grade 10 exams provide a preview of mandatory tests. Ottawa Citizen. p. A5.

Bosetti, L. (1996). Teacher evaluation as pedagogic opportunity: A case for portfolio assessment. The Canadian Administrator, 35, 8, 1-11.

Broadfoot, P. (1984). Selection, certification and control: Social issues in educational assessment. Philadelphia, PA: Falmer Press.

Bracey, G. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. Phi Delta Kappan, 68, 683-686.

Brooks, J., & Brooks, M. (1993). The case for constructivist classrooms. Alexandria, Va.: Association for Supervision and Curriculum Development.

Bankston, C., & Caldas, S. (1998). Race, poverty, family structure, and inequality in schools. Sociological Spectrum, 18, 55-75.

Caldas, S., & Bankston, C. (1997). The effect of school population socioeconomic status on individual student academic achievement. The Journal of Educational Research, 90, 269-277.

Calder, P. (1990). Impact of diploma examinations on the teaching-learning process. Edmonton: Alberta Teachers' Association.

Calder, P. (1990b). Impact of diploma examinations on the teaching-learning process. A study commissioned by The Alberta Teachers' Association.

Chalmers, R. (1999, March). Remedial help needed too. Edmonton Journal. p. A1.

Chwialkowska, L. (1999, April). Parents' education determines children's: Statscan. National Post. p. A6.

Cibulka, J., & Derlin, R. (1995). State educational performance reporting policies in the US: Accountability's many faces. International Journal of Educational Research, 23, 6, 479-492.

Cuttance, P. (1988). Assessing the effectiveness of schools. Edinburgh: Center for Educational Sociology. University of Edinburgh.

Conley, S., & Odden, A. (1992). Rethinking school finance: An agenda for the 1990s. San Francisco: Jossey-Bass.

Council on School Administration & the ATA. (1998). Accountability in education, teacher growth, supervision, and evaluation. Issues for Administration Series. Monograph 9. Edmonton.

Darling-Hammond, L. (1992). Reforming the school agenda. The School Administrator: Journal of the American Association of School Administrators, 49, 10, 22-27.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. Harvard Educational Review, 64, 1, 5-30.

Davis, S. (1998). Why do principals get fired? Principal, 78, 2, 34-39.

Dochy, F., Segers, M., & Buehl, M. (1999, April). A review of educational research. Principal, 96, 9, 14-18.

Donegan, M. (1998). Teacher and parent views on standardized testing: A cross-cultural comparison of the uses and influencing factors. Journal of Research in Childhood Education, 13, 1, 85-93.

Dougherty, E. (1998). What principals need to know about standards. Principal, 78, 2, 12-18.

Edmonds, R. (1984). A theory and design of social service reform. Social Policy, 15, 2, 57-64.

Earl, L. (1996). Comprehensive assessment and accountability. In K. Leithwood, The knowledge base for educational reform. (pp. 51-86). Toronto: University of Toronto, Ontario Institute for Studies in Education.

English, F. (1992). Deciding what to teach and test. Thousand Oaks, California: Corwin Press.

Evans, R. (1999, March). Money tied to marks in the education sector. Edmonton Sun. p. 28.

Fetler, M. (1991). Pitfalls of using SAT results to compare schools. American Educational Research Journal, 28, 2. 481-491.

Finn Jr., C. (1991). We must take charge of our schools and our future. New York, Toronto: The Free Press.

Fuhrman, S., & O'Day, J. (1996). Rewards and reform: Creating educational incentives that work. San Francisco: Jossey Bass.

Fullan, M. (1991). The new meaning of educational change. New York: Teachers' College Press.

Fullan, M. (1998). Breaking the bonds of dependency. Educational Leadership, 55, 7, 6-11.

Gall, D., Borg, R., & Gall, P. (1996). Educational research: An introduction (6 th ed.). New York: Longman Publishers USA.

Glickman, C. (1993). Renewing America's schools: A guide for school-based action. San Francisco: Jossey Bass.

Goldstein, H., & Cuttance, P. (1988). A note on national assessment and school comparisons. Journal of Education Policy, 3, 2, 197-202.

Gorsuch, R. (1983). Factor Analysis. Hillside, NJ.: Lawrence Erlabaum Associates.

Government announces new exams programs. (1983). ATA News, 17, 18,1.

Government of Alberta. (1999). Budget 99: The right balance. Edmonton: Alberta Education.

Gray, J., Jesson, D., Goldstein, H., et al. (1995). A multi-level analysis of school improvement: Changes in schools' performance over time. School Effectiveness and School Improvement, 6, 2, 97-114.

Gierl, M., Khaliq, S., & Boughton, K. (1999). Gender differential item functioning in mathematics and science: Prevalence and policy implications. Paper

presented at the Annual Meeting of the Canadian Society for the Study of Education, Sherbrooke, Quebec.

Guskey, T., & Kifer, E. (1990) Ranking school districts on the basis of statewide test results: Is it meaningful or misleading? Educational Measurement Issues and Practice, 9, 1, 11-16.

Haladyna, T., Haas, N., & Nolen, S. (1989). Test score pollution. Technical Report, 89, 1. Phoenix: Arizona State University West.

Haertel, E. (1986). The valid use of student performance measures for teacher evaluation. Educational Evaluation and Policy Analysis, 8, 1,45-60.

Haertel, E., & Walberg, H. (1993). Toward a knowledge base for school learning. Review of Educational Research, 63, 249-294.

Hart, C. (1988). The diploma dilemma. ATA Magazine, 68, 3, 11-13.

Henchey, N. (1998, May). Judging school quality. Montreal Gazette. p. 5.

Heneman, H. G. (1998). Assessment of the motivational reactions of teachers to a school-based performance awards program. Journal of Personnel Evaluation in Education, 12, 1, 43-59.

Herman, J. (1997). Large-scale assessment in support of school reform: Lessons in the search for alternative measure. CSE Technical Report 446. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Herman, J. (1999). The state of performance assessments. The School Administrator, 55, 110.

Hoffman, J. (1999). On the mark: Is standardized testing good for education? Today's Parent: Canada's Parenting Magazine, 11, 77-82.

Jencks, C. (1972). Inequity: A reassessment of the effect of family and schooling in America. New York: Basic Books.

Jeffs, A. (1999, March). School funding will depend on how well students perform. Edmonton Journal. p. A1.

Joint Advisory Committee. (1993). Principles for fair student assessment practices for education in Canada. Edmonton, Alberta: CRAME, University of Alberta.

Johnson, D., & Johnson, R. (1989). Cooperation and competition: Theory and research. Edina, Minn.: Interaction Books.

Jonassen, D., & Grabowski, B. (1993). Handbook of individual differences, learning, and instruction: Part VII prior knowledge. Hillsdale: Lawrence Erlbaum Associates.

Jones, M., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. Phi Delta Kappan, 11, 199-203.

Kane, M. (1998). Choosing between examinee-centered and test-centered standard-setting methods. Educational Assessment, 5, 3, 129-145.

Kelley, C. (1997). Teacher compensation and organization. Educational Evaluation and Policy Analysis, 19, 1,15-28.

Kelley, C., & Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. Educational Administration Quarterly, 33, 4, 474-505.

Kelley, C. (1999). The effects of organizational context on teacher expectancy. Educational Administration Quarterly, 41, 4, 44-55.

Kelley, C. (1998). The Kentucky school-based performance award program: School-level effects. Education Policy, 11, 75-83.

Kendall, L. (1995). Contextualization of school examination results 1992. Educational Research., 37, 2, 123-139.

Kim, J., & Mueller, C. (1978). Factor Analysis: Statistical Methods and Practical Issues. Beverly Hills, CA.: Sage Publications.

King, R., & Mathers, J. (1997). Improving schools through performance-based accountability and financial rewards. Journal of Education Finance, 23, 147-176.

King, R., & Verstegan, D. (1998). The relationship between school spending and student achievement: A review and analysis of 35 years of production function research. Journal of Education Finance, 24, 2, 243-262.

Kirst, M. (1990). Accountability: implications for state and local policymakers. Washington, DC: US Department of Education, Office of Educational Research and Improvement, Information Services.

Knight, A. (1993, June). Now's a perfect time to change the goofy way we pay teachers. The Denver Post, 1D, 5D.

Kohn, A. (1992). No contest: The case against competition (Rev. ed.). Boston, Mass.: Houghton Mifflin.

Kohn, A. (1998a). What to look for in a classroom and other essays. San Francisco: Jossey-Bass.

Kohn, A. (1998b). Only for my kid: How privileged parents undermine school reform. Phi Delta Kappan, 11, 156-161.

Kohn, A. (2000). The case against standardized testing: Raising the scores, ruining the schools. Portsmouth, NH: Heinemann.

Krembil, B. (1987). What do provincial exams measure? Alberta Science Teacher, 7, 1, 11-12.

Lawler, E. (1986). High involvement management. San Francisco: Jossey-Bass Publishers.

Lawler, E. (1990). Strategic pay. San Francisco: Jossey-Bass Publishers.

Levine, D. U., & Lezotte, L. (1990). Unusually effective schools. Madison, Wisconsin: National Center for Effective Schools Research and Development.

Linn, R. (1986). Educational testing and assessment: Research needs and policy issues. American Psychologist, 41, 1153-60.

Linn, R. (1998). Assessments and accountability. CSE Technical Report 490. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Linn, R., & Herman, J. (1997). Standards-led assessment: Technical and policy issues in measuring school and student progress. CSE Technical Report 426. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Lissitz, R. (1997). Statewide performance assessment: Continuity, context, and concerns. Contemporary Education, 69, 1, 15-19.

Loerke, D. (1993). Perceptions of the use of diploma examination results in teacher evaluation. Unpublished master's project, University of Alberta, Edmonton, Alberta.

Loerke, D. (1994). Perceptions of the use of diploma examination results in teacher evaluation. Applying Research to the Classroom, 12, 23-25. Department of Teacher Education and Supervision, Faculty of Education, University of Calgary.

Ludwig, J., & Bassi, L. (1999). The puzzling case of school resources and student achievement. Educational Evaluation and Policy Analysis, 21, 4, 385-403.

Macpherson, R. (1995). Accountability research in education: Current developments. International Journal of Educational Research, 23, 6, 561-567.

Madaus, G. (1981). Reactions to the Pittsburgh papers. Phi Delta Kappan, 62, 634-636.

Madaus, G. (1989). The distortion of teaching and testing: High-stakes testing and instruction. Peabody Journal of Education About Teachers and Teaching, 29, 46.

Maeroff, G. (1988). The empowerment of teachers: Overcoming the crisis of confidence. New York: Teachers College Press.

Marchesi, A. (1998). Meeting the needs of students at risk. Education Canada, 38, 2, 22-30.

Marshall, Andy. (1998, October). Stop making excuses: Education Minister Gary Mar wants public school board to take action to improve its test results. Calgary Herald, p. A17.

Mawhinney, H. (1995). Systematic reform, or new wine in an old bottle: Generating accountability through Ontario's new foundations in Education Reforms. International Journal of Educational Research, 23, 6, 505-518.

Mawhinney, H. (1998). Patterns of social control in assessment practices in Canadian frameworks for accountability in education. Educational Policy, 12, 98-109.

McConaghy, T. (1983, November). War declared on mediocrity. ATA Magazine, 67, 2, 10

McConaghy, T. (2000, December). The case against tougher standards. ATA Magazine, 81, 1, 20.

McEwen, N. (1998). Satisfaction with education in Alberta. Journal of Educational Research., 154, 1, 20-37.

McEwen, N. (1995). Accountability in education in Canada. Canadian Journal of Education, 20, 1, 1-7.

McEwen, N. (1998b). Selected findings and policy implications of large-scale assessments. Conference on Measurement and Evaluation: Current and Future Research Directions for the New Millennium. Banff, Alberta.

Meaghan, D., & Casas, F. (1995). On the testing of standards and standardized achievement testing: Panacea, placebo, or Pandora's box? Interchange, 26, 1, 33-58.

Mehrens, W. (1991). Defensible/indefensible instructional preparation for high stakes achievement teats: An exploratory trialog. A paper presented as part of a symposium with N. Cole and W. Popham at the annual meeting of the American Educational Research Association.

Mohrman, A. et al. (1994). School-based management: Organizing for high performance. San Francisco: Jossey-Bass.

Mohrman, A., Odden, R. & Mohrman, S. (1996). Aligning teacher compensation with systematic school reform: Skill-based pay and group-based performance awards. Educational Evaluation and Policy Analysis, 18, 1, 51-71.

Murname, R., & Levy, F. (1996). Teaching to new standards in Rewards and Reform, Fuhrman & O'Day. (pp. 78-112). San Francisco: Jossey-Bass.

Murphy, Joseph. (1998). Reforming America's schools 1980-2000. Journal of Educational Administration, 36, 5, 426-444.

Newmann, F., King, M., & Rigdon, M. (1997). Accountability and school performance: Implications for restructuring schools. Harvard Educational Review, 67, 1, 41-73.

Nickerson, R. (1989). New directions in educational assessment. Educational Researcher, 18, 9, 3-7.

No incentive for incentives program. (1999b). ATA News, 33, 19, 9.

Nolen, S., Haladyna, T., & Haas, N. (1992). Uses and abuses of achievement test scores. Educational Measurement: Issues and Practice, 11, 2, 9-15.

O'Shea, T., & Wideen, M. (1993). The impact of external examinations on science teaching. A paper presented at the National Association for Research in Science Teaching. Atlanta, GA.

Payne , L., & Michailides, M. (1998). New trends in large-scale assessment. Principal, 78, 2, 43-45.

Peters, F. (2000, February). Ten years later: Have reforms in education worked? ATA News, 43, 23, 2.

Peterson, B., & Neill, M. (1999). Alternatives to standardized tests. Rethinking Schools, 13, 3.

Phelps, Richard. (1998). The demand for standardized student testing. Educational Measurement: Issues and Practice, 17, 3, 5-23.

Phelps, R. (1999). Why testing experts hate testing. Fordham Report, 3, 1, 1-33.

Pipho, C. (1988). The uses of test scores. Phi Delta Kappan, 70, 4, 278-279.

Popham, W. (1987). The merits of measurement-driven instruction. Phi Delta Kappan, 68, 679-682.

Popham, J. (1998). Standardized tests should not be used to evaluate educational quality: A debate. Paper presented at the annual meeting of the American Educational Research Association. San Diego, CA.

Popham, W. (1999). Why standardized tests don't measure educational quality. Educational Leadership, 56, 6, 8-15.

Quality Counts '99. (1999, January). Education Week. 42-48.

Radenbush, S., & Bryk, A. (1987). Quantitative models for estimating teacher and school effectiveness. Paper presented at the Invitational Conference of the Center for Student Testing, Evaluation, and Standards. Princeton, New Jersey.

Ramirez, A. (1999). Assessment-driven reform: the emperor still has no clothes. Phi Delta Kappan, 11, 204-208.

Redfield, D. (1987). Expected student achievement as a potential factor for assessing teacher effectiveness. Paper presented at the Annual Meeting of the Mid-South Educational Research Association. Mobile, AL.

Redfield, D., & Craig, J. (1987). Identifying and documenting student outcomes for use in the evaluation of teachers: When standardized achievement tests do not apply. Paper presented at the Annual Meetings of the Mid-South Educational Research Association. Mobile, AL.

Raham, H. (1999). Linking assessment and school success. Paper presented to the American Educational Research Association. Montreal.

Reeves, D. (2000, March). Answers to common-sense questions about accountability. Thrust for Educational Leadership, 1,10-13.

Robertson, G. & Heyman, D. (1999, March). School board incentive raises ruckus. Calgary Herald. p. A8.

Rothman, R. (1995). Measuring up: Standards, assessment and school reform. San Francisco: Jossey-Bass.

Rothstein, R. (1998). Skewed comparisons. The School Administrator, 9, 20-24.

Samiroden, W. (1987). Diploma examinations: Limiting what is taught. ATA Magazine, 67, 3, 4-7.

Samiroden, W. (1988, March). Diploma examinations: Some effects on teaching and learning. Paper presented at Calgary Teachers' Convention. Calgary, Alberta.

Samiroden, W. (1991). Diploma examinations in science: Their effects on teaching and learning. Alberta Science Education Journal, 24, 1, 8-23.

Schafer, W., & Lissitz, R. (1987). Measurement training for school personnel: Recommendations and reality. Journal of Teacher Education, 38, 3, 57-63.

Schmoker, M. (1996). Results: The key to continuous improvement. Alexandria. Va.: Association for Supervision and Curriculum Development.

Schmoker, M. (1997) Setting goals in turbulent times. In Rethinking Educational Change with Heart and Mind: 1997 ASCD Yearbook. (pp. 128-147). Alexandria, Va.: Association for Supervision and Curriculum Development.

Shepard, L. (1989). Why we need better assessments. Educational Leadership, 46, 7, 4-9.

Shepard, L. (1990, Fall). Inflated test scores gains: Is the problem old norms or teaching the test?. Educational Measurement: Issues and Practice, 38, 15-21.

Sirotnik, K., & Kimball, K. (1999). Standards for standards-based accountability systems. Phi Delta Kappan, 11, 209-214.

Smith, M. (1990). Put to the test: The effects of external testing on teachers. Educational Researcher, 20, 8-11.

Snow, R. (1989). Toward assessment of cognitive and conative structures in learning. Educational Researcher, 18, 9, 8-14.

Stake, R. (1993). Differentiation and standardization of aims. Paper presented at the annual general meeting of the American Educational Research Association.

Stake, R. (1998). Some comments on assessment in U.S. education, Education Policy Analysis Archives, 6, 14, 1-7.

Stake, R. (1998). Some comments on assessment in US education. Educational Policy Analysis, 6, 14.

Stiggins, R. (1989). A commentary on the role of student achievement data in evaluation of teachers. Journal of Personnel Evaluation in Education, 3, 7-15.

Stiggins, R. (1991). Assessment literacy. Phi Delta Kappan, 72, 534-539.

Stiggins, R. (1999a). Assessment, student confidence, and school success. Phi Delta Kappan, 11, 191-198.

Stiggins, R. (1999b, March). Are you assessment literate? The High School Magazine, 20-23.

Stiggins, R., & Conklin, N. (1992). In teachers' hands: Investigating the practices of classroom assessment. New York: SUNY Press.

Swanson, J. (1994). The impact of diploma examinations on the teaching-learning process: A concept map. Unpublished thesis, University of Alberta, Edmonton, Alberta, Canada.

Sykes, G. (1997). On trial: The Dallas value-added accountability system. In Millman, J. Grading teachers, grading schools: Is student achievement a valid evaluation measure? (pp. 110-119). Thousand Oaks, Ca: Corwin Press.

Task force reports public wants exams. (1983). ATA News, 17, 15, 8.

Teachers see Klein's ministry changes as mixed bag. (1999a). ATA News, 33,19, 3.

Top marks: The demise of learning has been greatly exaggerated. (1998, October). Calgary Herald. p. A22.

Wasserman, D. (1999). Comparing diploma examination results over time. Student Evaluation Branch. Edmonton: Alberta Learning.

Watts, J., Gaines, G., & Creech, J. (1998). Getting results: A fresh look at school accountability. Atlanta: Southern Regional Education Board.

Wang, M., Haertel, G., & Walberg, H. J. (1993). What influences learning? A content analysis of review literature. Journal of Educational Research, 84, 30-43.

Wideen, M., O'Shea, T., & Ivany, G. (1992). External examinations and the curriculum: Do they monitor or control? Paper presented at the annual meeting of the American Research Association. San Francisco.

Wiggins, G. (1994). None of the above. The Executive Educator, 16 , 7, 14-18.

U.S. General Accounting Office. (1993). Student testing: Current extent and expenditures, with cost estimates for national examination (GAO/PEMD-93-9-8). Washington, DC: General Accounting Office, Program Evaluation and Methodology Division.

U.S. Office of Technology Assessment. (1992). Testing in American schools: Asking the right questions (OTA-SET-519). Washington, DC: Author.

Unland, Karen. (1999, May). Teachers won't buy into incentive program. Edmonton Journal. p. A1.

Unland, Karen. (2000, July). School rankings pressure teachers. Edmonton Journal. p. B3.

Yin, R. K. (1989). Case study research: Design and methods (Rev.ed.). Newbury Park, CA: Sage.

# APPENDIX A

# QUESTIONNAIRE

School Code: S/M/L (V1)

**Please fill in <u>all</u> the circles that apply to you**

I am willing to participate in this study O Yes O No (V2)
**If you select yes, please continue.**
**If you selected no, please return this questionnaire unanswered.**

I participated in a similar study in 1994 O Yes O No (V3)
I have taught a diploma subject O Yes O No (V4)

1) My roles in this school include
O principal (V5) O vice principal (V6) O department head (V7)
O teacher (V8) O librarian (V9) O counselor (V10)

2) My instructional role(s) include
O English (V11) O Chemistry (V15) O CTS (V19)
O Social Studies (V12) O Physics (V16) O Physical Ed. (V20)
O Languages (V13) O Biology (V17) O Art (V21)
O Mathematics (V14) O Music (V18) O Other (V22)
O not applicable (V23)

3) The course level(s) I teach at include
O 10 (V24) O 13 (V27) O other (V30)
O 20 (V25) O 23 (V28)
O 30 (V26) O 33 (V29)

4) Including this year, I have taught for
O 1-2 years (V31) O 6-10 years (V33) O 21+ years (V35)
O 3-5 years (V32) O 11-20 years (V34)

5) Including this year, I have taught in my current position for
O 1-2 years (V36) O 6-10 years (V38) O 21+ years (V40)
O 3-5 years (V37) O 11-20 years (V39)

6) My educational background includes
O B. Education (V41) O M. Education (V44) O Ph.D. Education (V47)
O B. Science (V42) O M. Science (V45) O Ph.D. Science (V48)
O B. Arts (V43) O M. Arts (V46) O Ph.D. Arts (V49)
O Other _____(please specify) (V50)

7) I have a specialization/degree in my primary teaching assignment (V51)
O Yes _____ (please specify) O No Onot applicable

8) I have a specialization/degree in the position I hold (e.g. Ed. Admin. (V52)
O Yes _____ (please specify) O No Onot applicable

| **Use this key to answer questions 9 to 50** | Know to be true | Believe to be true | Do not know/Unsure | Believe to be false | Know to be false | |
|---|---|---|---|---|---|---|
| 9) I have read the Alberta Learning published guidelines for interpreting diploma examination results. | O | O | O | O | O | (V53) |
| 10) This school's administration are aware of the published guidelines and resources available for interpreting diploma examination results. | O | O | O | O | O | (V54) |
| 11) This district's administration are aware of the published guidelines and resources available for interpreting diploma examination results. | O | O | O | O | O | (V55) |
| 12) Diploma teachers use the Examiners' Reports and Jurisdiction Reports to analyze diploma examination results. | O | O | O | O | O | (V56) |
| 13) Internet access to diploma examination information by administration, teachers, and the public has had a positive effect on diploma results. | O | O | O | O | O | (V57) |
| 14) The administration share the details of the school 's diploma results with all teaching staff. | O | O | O | O | O | (V58) |
| 15) The diploma examination program has been instrumental in aligning high school curriculum in Alberta. | O | O | O | O | O | (V59) |
| 16) Diploma examination results are used fairly in my/this school. | O | O | O | O | O | (V60) |
| 17) The publishing of individual school results by the media has had little or no effect on the way in which diploma results are used at this school. | O | O | O | O | O | (V61) |
| 18) All teachers at this school are recognized for their contribution toward desirable diploma results. | O | O | O | O | O | (V62) |
| 19) Diploma teachers at this school are given various rewards and perks in recognition of good diploma results. | O | O | O | O | O | (V63) |

| | | | | | | |
|---|---|---|---|---|---|---|
| 20) Mandatory provincial diploma examinations have had a positive effect on student academic performance. | O | O | O | O | O | (V64) |
| 21) Teachers have been provided with additional support resources and/or mentors to improve diploma results. | O | O | O | O | O | (V65) |
| 22) Diploma examination results are used *informally* by this school's administration as part of teacher evaluation. | O | O | O | O | O | (V66) |
| 23) Diploma examination results are used *formally* by this school's administration as part of teacher evaluation. | O | O | O | O | O | (V67) |
| 24) Diploma examination results are used *informally* by this district's administration as part of teacher evaluation. | O | O | O | O | O | (V68) |
| 25) Diploma examination results are used *formally* by this district's administration as part of teacher evaluation. | O | O | O | O | O | (V69) |
| 26) Diploma examination results are used *informally* by this district's administration to as part of principal evaluation. | O | O | O | O | O | (V70) |
| 27) Diploma examination results are used *formally* by this district's administration to as part of principal evaluation. | O | O | O | O | O | (V71) |
| 28) Diploma examination results are used by this school's administration to determine teaching assignments. | O | O | O | O | O | (V72) |
| 29) Classroom instruction has improved as a results of the re-introduction of diploma examinations in Alberta. | O | O | O | O | O | (V73) |
| 30) Teachers have been transferred or have asked for transfers because of poor diploma results. | O | O | O | O | O | (V74) |
| 31) Teachers have been promoted to leadership roles because of good diploma results. | O | O | O | O | O | (V75) |
| 32) There is a greater level of stress associated with teaching a diploma subject than a non-diploma subject. | O | O | O | O | O | (V76) |
| 33) Diploma teachers are as willing as non-diploma teachers to implement new curricula and/or teaching methods. | O | O | O | O | O | (V77) |
| 34) Teachers at this school have asked for non-diploma teaching assignments because of poor diploma results. | O | O | O | O | O | (V78) |

| | | | | | | |
|---|---|---|---|---|---|---|
| 35) Teachers at this school are encouraged to include diploma results in their professional growth plans. | O | O | O | O | O | (V79) |
| 36) The diploma examination program has reduced the professional autonomy of teachers. | O | O | O | O | O | (V80) |
| 37) Temporary or interim teachers with good diploma results are more likely to have their contracts renewed than those with lower than provincial averages. | O | O | O | O | O | (V81) |
| 38) The use of diploma results in teacher assessment has increased in the past 5 years. | O | O | O | O | O | (V82) |
| 39) At this school, efforts to improve student learning are most often focused on ways to increase diploma results. | O | O | O | O | O | (V83) |
| 40) Since the introduction of 33 level diploma examinations, there has been an increase in the percentage of students at this school taking 33 over 30 level courses. | O | O | O | O | O | (V84) |
| 41) Positive recognition is given to those teachers whose school awarded marks closely align with diploma marks. | O | O | O | O | O | (V85) |
| 42) Students at this school are encouraged to complete 30 level courses even if they are having difficulty. | O | O | O | O | O | (V86) |
| 43) At this school the grade point entrance requirements for 30 level academic courses is greater than 50%. | O | O | O | O | O | (V87) |
| 44) Teachers encourage students to select particular course based on the student's anticipated success on the diploma. | O | O | O | O | O | (V88) |
| 45) At this school, students select subjects based on their anticipated success on the diploma examinations. | O | O | O | O | O | (V89) |
| 46) Diploma results are an important criterion used by students/parents when selecting a high school. | O | O | O | O | O | (V90) |
| 47) Post-secondary institutions give preference to students with 30 level course over those with 33 level course. | O | O | O | O | O | (V91) |
| 48) The media has had a positive effect on the appropriate use of diploma examination results. | O | O | O | O | O | (V92) |

| | | | | | | |
|---|---|---|---|---|---|---|
| 49) The implementation of diplomas has resulted in the improvement of student assessment practices by teachers. | O | O | O | O | O | (V93) |
| 50) Teachers/administrators receive adequate professional training in the assessment of student learning. | O | O | O | O | O | (V94) |

***If you wish to include any comments regarding specific items on this questionnaire, please use the space provided on the following page.***

**Question Number** **Comment**

***Thank you for your participation in this research study.***

**Please return questionnaire to the designated individual at your school or directly to:**

**Marlene McDonald**
**c/o Margaret Haughey**
**Professor, & Associate Chair**
**7-104 Education North**
**Department of Policy Studies,**
**University of Alberta, AB**
**T6G 2G5**

# APPENDIX B

# STUDY CORRESPONDENCE

# Letter to Superintendents

December 1999

Dear _____

I am requesting permission to conduct a survey at randomly selected schools in your jurisdiction. The survey is part of the research base of my master's thesis in the Department of Education Policy Studies at the University of Alberta and has passed the Faculty of Education's research ethics review.

My study is on the "Role of Diploma Examinations in Alberta High Schools." The attached survey questionnaire is intended to gauge the attitudes and beliefs of school based administrators and high school teaching staff on the use of diploma examination results. The purpose of this study is to investigate how diploma examination results are being used and the extent that these uses have changed since the 1984 re-implementation of the diploma program in Alberta.

I am requesting permission to contact the following school(s) in your jurisdiction to participate in a survey questionnaire:
**List of Schools**

In addition to participation in the questionnaire, individual teachers or administrators may be selected from your jurisdiction for semi-standardized interviews. The commitment for each participant would be a single interview that should take a maximum of 1 hour to complete. A stratified sample of interview candidates will be selected to represent various stakeholders. The list of interview questions is available upon request. In order to protect interview participants' confidentiality, their identity will only be known to the researcher and interviews will be conducted off school property.

A copy of the questionnaire is enclosed for your records. The questionnaire should require approximately 30-45 minutes to complete and the administration date/time will be at the discretion of the principal. Each questionnaire is identified by a school code/school size for purposes of tracking and matching responses. This code is be known only to the researcher. The coded data collected will be destroyed upon completion of the analysis. All responses will be held in complete confidence as results of this study will be reported in general terms with no school based reporting nor individual teacher disclosures undertaken. Copies of the research results will be made available at the completion of this study.

If you have any concerns or questions about the questionnaire, the interviews, or on confidentiality, please contact me at (780) 436-7204 (home) or Margaret Haughey at (780) 492-7609 (University of Alberta).

A reply from you would be most appreciated as soon as possible. I hope to administer the questionnaire during the month of February and conduct interviews shortly after. I have taken the liberty of including the following standard reply form.

Thank you for your co-operation in this endeavor.

Sincerely,

Marlene McDonald
Department of Policy Studies
University of Alberta

---

Permission is (granted/denied) to Marlene McDonald to conduct a survey of school based administration and high school teachers on the "Role of Diploma Examinations in Alberta High Schools" in this school jurisdiction.

_____ _____

Superintendent Date

_____

Jurisdiction

Please mail the completed form to:

Marlene McDonald
c/o Margaret Haughey
7-104 Education North
Department of Policy Studies
University of Alberta
T6G 2G5

# Letter to Principals

April 5, 2000

To: School Principals
Re: Role of Diploma Examination Results Research Survey

I am requesting permission to conduct a survey at randomly selected schools that offer grade 12 diploma subjects in your jurisdiction. Your superintendent has granted permission in your jurisdiction. The survey is part of the research base for my master's thesis in the department of Policy Studies at the University of Alberta. The study has received approval from the Faculty of Education's research ethics review.

My study is on the "Role of Diploma Examinations in Alberta High Schools." The attached survey questionnaire is intended to gauge the attitudes and beliefs of school based administrators and high school teaching staff on the use of diploma examination results. The purpose of this study is to investigate how diploma examination results are being used and the extent that these uses have changed since the 1984 re-implementation of the diploma program in Alberta.

I am requesting permission to conduct this survey of the senior high teaching staff and administration at your school. The questionnaire should take approximately 1/2 hour to complete and the administration date/time will be at the discretion of the principal. Each questionnaire is identified by a school code/school size for purposes of tracking and matching responses. All responses will be held in complete confidence, as results of this study will be reported in general terms with no school based reporting nor individual teacher disclosures undertaken. Copies of the research results will be made available at the completion of this study.

In addition to participation in the questionnaire, individual teachers or administrators may be selected from your jurisdiction and invited to participate in a semi-standardized interview. A stratified sample of twelve interview candidates will be selected to represent various stakeholders from all the stakeholder sin the survey. The commitment for each participant would be a single interview that should take a maximum of one hour to complete. The list of interview questions is available upon request. In order to protect interview participants' confidentiality, their identity will only be known to the researcher and interviews will be conducted off school property.

A reply from you as soon as possible would be most appreciated as I hope to have the questionnaires administered before the end of April and to conduct the interviews shortly thereafter. Enclosed along with this permission letter is a copy of the questionnaire, a copy of the ethics review document form the University of Alberta, and a standard reply form. I am required to have written permission form each principal before I can start this research study.

If you or your staff have any concerns about the research study, the questionnaire, the interviews, or on confidentiality, please contact me by telephone at (780) 436-7204 (home) or Margaret Haughey at (780) 492-7609 (University of Alberta).

Thank you for your co-operation in this endeavor.

Sincerely,

Marlene McDonald
Department of Policy Studies
University of Alberta

Margaret Haughey, Ph.D.
Professor, & Associate Chair
Educational: Policy Studies
7-104 Education North
Department of Policy Studies
University of Alberta, AB
T6G 2G5

Tel: 780/492-7609
Fax: 780/492-2024
margaret.haughey@ualberta.ca

*Please return the completed reply form using the postage paid envelope provided.*

---

Permission is (granted/denied) to Marlene McDonald to conduct a survey of school-based administration and high school teachers on the "Role of Diploma Examinations in Alberta High Schools" at this school.

_____

Principal (or representative) Date

_____

School Jurisdiction

Anticipated number of copies of the survey questionnaire required _____.

# Letter to Participants

April 5, 2000

To: All participants
Re: Attached survey questionnaire

The attached survey questionnaire is intended to gauge the attitudes and beliefs of school based administrators and high school teaching staff on the use of diploma examination results. The purpose of this study is to investigate the how diploma examination results are being used and to what extent these uses have changed since the re-implementation of the Alberta diploma program in 1984.

The questionnaire should require approximately 1/2 hour to complete. Each questionnaire is identified by a school code/school size for purposes of tracking and matching responses. No names should be placed on the questionnaire in order to ensure each respondent's anonymity is respected. The school code/size data collected will be destroyed upon completion of the analysis. All responses will be held in complete confidence as results of this study will be reported in general terms with no school based reporting nor individual teacher disclosures undertaken. Copies of the research results will be made available to the superintendents of the participating jurisdictions at the completion of this study.

Your participation in this survey is important. You are contributing to research into how student diploma marks are used. With changes in teacher evaluation practices, increased emphasis on accountability, FOIPP, and numerous other changes in education policy, it is important to monitor how results are being used and to remain vigilant about potential misuse. The first step in this monitoring process is to gauge the attitudes and beliefs of high school teachers and administration. This study has the potential of affecting assessment practices in Alberta.

I would greatly appreciate if you would participate in this study by completing the attached questionnaire. It is important that teacher assessment practices in Alberta are fair and grounded in education literature. Regardless of your decision on participation, please seal your completed/uncompleted questionnaire in the attached envelope. You may elect to return this information directly to me or to the individual in your school charged with the collection and forwarding of survey questionnaires. Questionnaire may be returned to the address designated below.

If you or your staff have any concerns about the research study, the questionnaire, or on confidentiality, please contact me at (780) 436-7204 (home) or Margaret Haughey at (780) 492-7609 (University of Alberta).

Thank you for your co-operation in this endeavor.

Marlene McDonald

Questionnaires may be returned by mail to:

Marlene McDonald
c/o Margaret Haughey
Professor, & Associate Chair
7-104 Education North
Department of Policy Studies
University of Alberta, AB
T6G 2G5

# APPENDIX C

# INTERVIEW AND CORRESPONDENCE

Interview Participation Letter

May 2000

To: _____

I am a student at the University of Alberta working on my master's thesis in Education Policy Studies. The topic of my study is "The Role of Diploma Examination Results in Alberta High Schools." I am using two data collection methods as part of my research. I am conducting a questionnaire survey of randomly selected Alberta high schools along with a stratified set of interviews involving various stakeholder representatives. The interview is intended is intended to gauge the attitudes and beliefs of school-based administrators and high school teaching staff on the use of diploma examination results. The purpose of this study is to investigate how diploma examination results are being used and to what extent these uses have changed since the re-implementation of the Alberta diploma program in 1984.

The interviews are designed with questions that complement those of the questionnaire. The interview method allows participants to elaborate on specific items and to help clarify emerging issues for the researcher. I have selected you as a potential interview candidate to represent one of the stakeholder groups in this study.

Your superintendent has given me permission to conduct this study in your jurisdiction with the understanding that the identity of the interviewees will only be known to the researcher. Participation in this interview is voluntary and the interview will be conducted off school property. All responses will be held in complete confidence, as results of this study will be reported in general terms with no school based reporting nor individual teacher disclosures undertaken. Participants' consent forms, identifying records, and the information collected during the interview will be secured by the researcher to ensure confidentiality of the respondents. all information that could be used to identify an individual or school will be known only to the researcher and destroyed when the study is competed. The results of this study will only be used to complete my master's thesis. An executive summary of the research results will be forwarded to each participating school jurisdiction upon completion of this study.

The commitment for each participant would be a single interview that should take a maximum of 1 hour to complete. I am hoping that the interviews will corroborate and extend the information gathered from the questionnaire. A copy of the survey questionnaire is available upon request. If you decide to participate, you may exercise your right to opt out at any time during the interview process or may elect to answer only particular questions. I am also requesting permission to audio tape the interview on the consent form. More accurate notes of the interview are possible with a taped response; however, the

use of audio tape is at the discretion of each participant. You will be provided with a copy of this letter and the signed consent form for your records.

Your participation in this survey is important. You are contributing to research into how student marks are used. With changes in teacher evaluation practices, increased emphasis on accountability, FOIPP, and numerous other changes in education policy, it is important to monitor how results are being used and to remain vigilant about potential misuse. The first step in this monitoring process is to gauge the attitude and beliefs of high school teachers and administration. This study has the potential of affecting teacher evaluation practices in Alberta. The results of this study will be available to participants upon completion of the data analysis.

I would greatly appreciate if you would participate in this study. It is important that diploma examination results in Alberta are used appropriately and that their use is grounded in educational literature. Please complete the following consent form to indicate your willingness to participate in this study.

If you have any concerns or questions about the interview, anonymity, or on confidentiality, please contact me at (780) 436-7204 (home) or Margaret Haughey at (780) 492-7609 (University of Alberta).

Thank you for your anticipated consideration of this enterprise,

Marlene McDonald
Department of Policy Studies
University of Alberta

---

**Interview Consent Form**

I have read the attached interview participation letter and I am willing to participate in a single interview for the completion of Marlene McDonald's master's thesis on the "Role of Diploma Examinations in Alberta High Schools."

My signature indicates that I understand the nature of this study and that I am willing to be interviewed. The interview format, time, and location will be mutually agreed upon.

| _____ | _____ | _____ |
|---|---|---|
| Name (Please Print) | Professional Role | School |
| _____ | _____ | |
| Signature | Date | |

Permission is (granted/denied) to Marlene McDonald for the audio taping of the interview. _____ (Please Initial)

**Interview Format**

1) Review of Interview request letter information.

2) Clarification of interview content/use/concerns.

3) Reminder of participants right to opt out or to answer only select questions

4) How the interview information will be recorded and the optional use of audio tape.

5) Overview of the 5 parts of this interview and estimated time frame:

The first part of the interview starts with some biographical questions that are followed by questions in four related categories:

i) How well teachers are informed?

ii) School/district policy regarding diploma courses

iii) The use of student results

iv) Effect of the media

6) The interview begins with a recording of a confirmation of the participant's understanding of the nature of the study and their willingness to participate in the interview.

## Interview Questions

**Part 1 - Biographical Questions**
1) Did you participate in a similar study conducted in 1993?
2) What are your professional roles at this school?
3) What do you teach? At what levels?
4) How long have you taught (including this year)?
5) How long have you been at your current position?
6) Have you taught a diploma subject? When?
7) May I ask you what your educational background is?
8) Do you have a specialization/degree in your primary school assignment?

**General Question:**
What is the role of a diploma examination?
Comment on the strength and weakness of the current examination program.

**Part 2 - How well are teachers informed?**
1) How familiar are you with the published guidelines for interpreting diploma exam results from Alberta Learning formerly Alberta Education ?

2) To what extent do you think your school administration are aware of the these guidelines and/or other resources available for interpreting diploma examination results? How about the district administration? Should they be?

3) Do you know if the diploma teachers at your school use the Examiners' Reports and Jurisdiction Reports to analyze diploma examination results? Any idea how they use these reports?

4) Are you aware of the Alberta Learning web site on the Internet? What effect has Internet access to diploma examination information had?

5) How well informed are you regarding your schools diploma examination results? How does the administration share the details of your school diploma results with the teaching staff?

**Part 3 - School/district policy regarding diploma courses**

1) What is the grade point entrance requirement for 30/33 level courses at your school? In your district? What do you think it should be?

2) Who decides which students should write diploma examinations?

3) To what extent do efforts to improve student learning, such as professional development activities, focus on ways to increase diploma examination marks?

4) In the past 5 years, there has been a change in the percentage of students at some schools taking 33 level courses over 30 level courses? Has there been a change in enrollment in Science 30? Can you explain these trends?

5) If diploma examinations were no longer mandatory, provincial assessment ceased to exist, how do you think student performance would be affected?

6) Is there a connection between good results and good teaching?

**Part 4 - The use of student results**

1) In your opinion has the use of diploma results in teacher assessment changed in the past 5 years? In what ways?

2) Are teachers at this school encouraged to include diploma results in their professional growth plans?

3) To what extent are diploma examination results used by your school's administration to determine teaching assignments? To select teachers?

4) In your experience have teachers been "promoted" to leadership roles because of good diploma results?

5) Why do you think some teachers request non-diploma teaching assignments?

6) What is your estimate of stress associated with teaching a diploma subject compared to that of a non diploma subject?

7) Are teachers provided with additional support resources or mentors to improve diploma results? Reduced class sizes?

8) In what ways are diploma teachers recognized at your school for good diploma results? Are all the teachers recognized for their contribution toward desirable diploma results?

9) Have teachers ever been transferred by the school's/or district's administration because of poor diploma results? Have they asked for transfers?

10) To what extent are student results used informally or formally by this school's administration in teacher evaluation? Principal evaluation? How do you think the district's administration uses test results?

**Part 5 - Effect of the media**

1) To what extent are students or their parents selecting high schools based on diploma results?

2) Do students in your school select their diploma subject teachers based on past diploma results?

3) If the media stopped publishing individual school results in newspapers and on the Internet, what would change? What effect has the media had on the use of diploma results?

**Final Comments:**

Any additional comments regarding the role of diploma examinations that you would like to add?

# APPENDIX D

# STATISTICAL TABLES

**Appendix 1:**

**Individual Item Analysis by Variable Frequency, Percent, Mean, and S. D.**

| Variable | Response | Frequency | Valid % | Mean | S.D. |
|---|---|---|---|---|---|
| 53 | Know to be true | 194 | 51.2 | | |
| | Believe to be true | 68 | 17.9 | | |
| | Do not know/unsure | 43 | 11.3 | | |
| | Believe to be false | 12 | 3.2 | | |
| | Know to be false | 62 | 16.4 | | |
| | Total | 379 | 100.0 | 2.16 | 1.49 |
| 54 | Know to be true | 218 | 56.2 | | |
| | Believe to be true | 125 | 32.2 | | |
| | Do not know/unsure | 43 | 11.1 | | |
| | Believe to be false | 2 | 0.5 | | |
| | Know to be false | 0 | 0 | | |
| | Total | 388 | 100.0 | 1.56 | 0.71 |
| 55 | Know to be true | 173 | 44.7 | | |
| | Believe to be true | 162 | 41.9 | | |
| | Do not know/unsure | 52 | 13.4 | | |
| | Believe to be false | 0 | 0 | | |
| | Know to be false | 0 | 0 | | |
| | Total | 387 | 100.0 | 1.69 | 0.70 |
| 56 | Know to be true | 207 | 53.5 | | |
| | Believe to be true | 120 | 31.0 | | |
| | Do not know/unsure | 53 | 13.7 | | |
| | Believe to be false | 4 | 1.0 | | |
| | Know to be false | 3 | .8 | | |
| | Total | 387 | 100.0 | 1.65 | 0.81 |
| 57 | Know to be true | 21 | 5.4 | | |
| | Believe to be true | 66 | 17.1 | | |
| | Do not know | 255 | 66.1 | | |
| | Believe to be false | 37 | 9.6 | | |
| | Know to be false | 7 | 1.8 | | |
| | Total | 386 | 100.0 | 2.85 | 0.73 |
| 58 | Know to be true | 269 | 69.9 | | |
| | Believe to be true | 59 | 15.3 | | |
| | Do not know/unsure | 29 | 7.5 | | |
| | Believe to be false | 19 | 4.9 | | |
| | Know to be false | 9 | 2.3 | | |
| | Total | 385 | 100.0 | 1.55 | 0.99 |

Continued

**Appendix 1: Individual Item Analysis (Continued)**

| Variable | Response | Frequency | Valid % | Mean | S.D. |
|---|---|---|---|---|---|
| 59 | Know to be true | 72 | 18.6 | | |
| | Believe to be true | 172 | 44.4 | | |
| | Do not know/unsure | 121 | 31.3 | | |
| | Believe to be false | 18 | 4.7 | | |
| | Know to be false | 4 | 1.0 | | |
| | Total | 387 | 100.0 | 2.25 | 0.85 |
| 60 | Know to be true | 130 | 33.9 | | |
| | Believe to be true | 170 | 44.3 | | |
| | Do not know/unsure | 69 | 18.0 | | |
| | Believe to be false | 11 | 2.9 | | |
| | Know to be false | 4 | 1.0 | | |
| | Total | 384 | 100.0 | 1.93 | 0.85 |
| 61 | Know to be true | 58 | 15.0 | | |
| | Believe to be true | 111 | 28.8 | | |
| | Do not know/unsure | 92 | 23.8 | | |
| | Believe to be false | 97 | 25.1 | | |
| | Know to be false | 28 | 7.3 | | |
| | Total | 386 | 100.0 | 2.81 | 1.18 |
| 62 | Know to be true | 79 | 20.4 | | |
| | Believe to be true | 130 | 33.5 | | |
| | Do not know/unsure | 90 | 23.2 | | |
| | Believe to be false | 70 | 18.0 | | |
| | Know to be false | 19 | 4.9 | | |
| | Total | 388 | 100.0 | 2.54 | 1.15 |
| 63 | Know to be true | 2 | 0.5 | | |
| | Believe to be true | 6 | 1.5 | | |
| | Do not know/unsure | 65 | 16.8 | | |
| | Believe to be false | 136 | 35.1 | | |
| | Know to be false | 179 | 46.1 | | |
| | Total | 388 | 100.0 | 4.25 | 0.82 |
| 64 | Know to be true | 25 | 6.4 | | |
| | Believe to be true | 128 | 33.0 | | |
| | Do not know/unsure | 149 | 38.4 | | |
| | Believe to be false | 71 | 18.3 | | |
| | Know to be false | 15 | 3.9 | | |
| | Total | 388 | 100.0 | 2.81 | 0.94 |

Continued

**Appendix 1: Individual Item Analysis (Continued)**

| Variable | Response | Frequency | Valid % | Mean | S.D. |
|---|---|---|---|---|---|
| 65 | Know to be true | 27 | 7.0 | | |
| | Believe to be true | 88 | 22.7 | | |
| | Do not know/unsure | 92 | 23.7 | | |
| | Believe to be false | 122 | 31.4 | | |
| | Know to be false | 59 | 15.2 | | |
| | Total | 388 | 100.0 | 3.25 | 1.70 |
| 66 | Know to be true | 17 | 4.4 | | |
| | Believe to be true | 70 | 18.2 | | |
| | Do not know/unsure | 161 | 41.8 | | |
| | Believe to be false | 92 | 23.9 | | |
| | Know to be false | 45 | 11.7 | | |
| | Total | 385 | 100.0 | 3.20 | 1.01 |
| 67 | Know to be true | 7 | 1.8 | | |
| | Believe to be true | 26 | 6.5 | | |
| | Do not know/unsure | 134 | 33.7 | | |
| | Believe to be false | 150 | 37.7 | | |
| | Know to be false | 81 | 20.4 | | |
| | Total | 398 | 100.0 | 3.68 | 0.93 |
| 68 | Know to be true | 12 | 3.0 | | |
| | Believe to be true | 69 | 17.2 | | |
| | Do not know/unsure | 215 | 53.6 | | |
| | Believe to be false | 87 | 21.7 | | |
| | Know to be false | 18 | 4.5 | | |
| | Total | 401 | 100.0 | 3.07 | 0.83 |
| 69 | Know to be true | 7 | 1.7 | | |
| | Believe to be true | 32 | 8.0 | | |
| | Do not know/unsure | 213 | 53.0 | | |
| | Believe to be false | 112 | 27.9 | | |
| | Know to be false | 38 | 9.5 | | |
| | Total | 402 | 100.0 | 3.35 | 0.83 |
| 70 | Know to be true | 8 | 2.0 | | |
| | Believe to be true | 74 | 18.4 | | |
| | Do not know/unsure | 229 | 57.0 | | |
| | Believe to be false | 81 | 20.1 | | |
| | Know to be false | 10 | 2.5 | | |
| | Total | 402 | 100.0 | 3.03 | 0.75 |

Continued

**Appendix 1: Individual Item Analysis (Continued)**

| Variable | Response | Frequency | Valid % | Mean | S.D. |
|---|---|---|---|---|---|
| 71 | Know to be true | 6 | 1.5 | | |
| | Believe to be true | 29 | 7.2 | | |
| | Do not know/unsure | 256 | 63.7 | | |
| | Believe to be false | 96 | 23.9 | | |
| | Know to be false | 15 | 3.7 | | |
| | Total | 402 | 100.0 | 3.21 | 0.69 |
| 72 | Know to be true | 32 | 7.9 | | |
| | Believe to be true | 102 | 25.3 | | |
| | Do not know/unsure | 167 | 41.4 | | |
| | Believe to be false | 78 | 19.4 | | |
| | Know to be false | 24 | 6.0 | | |
| | Total | 403 | 100.0 | 2.90 | 1.00 |
| 73 | Know to be true | 15 | 3.7 | | |
| | Believe to be true | 114 | 28.3 | | |
| | Do not know/unsure | 157 | 39.0 | | |
| | Believe to be false | 95 | 23.6 | | |
| | Know to be false | 22 | 5.5 | | |
| | Total | 403 | 100.0 | 2.99 | 0.94 |
| 74 | Know to be true | 11 | 2.7 | | |
| | Believe to be true | 38 | 9.5 | | |
| | Do not know/unsure | 236 | 58.7 | | |
| | Believe to be false | 99 | 24.6 | | |
| | Know to be false | 18 | 4.5 | | |
| | Total | 402 | 100.0 | 3.19 | 0.77 |
| 75 | Know to be true | 7 | 1.7 | | |
| | Believe to be true | 53 | 13.2 | | |
| | Do not know/unsure | 231 | 57.3 | | |
| | Believe to be false | 90 | 22.3 | | |
| | Know to be false | 22 | 5.5 | | |
| | Total | 403 | 100.0 | 3.17 | 0.79 |
| 76 | Know to be true | 200 | 49.8 | | |
| | Believe to be true | 146 | 36.3 | | |
| | Do not know/unsure | 28 | 7.0 | | |
| | Believe to be false | 24 | 6.0 | | |
| | Know to be false | 4 | 1.0 | | |
| | Total | 402 | 100.0 | 1.72 | 0.91 |

Continued

**Appendix 1: Individual Item Analysis (Continued)**

| Variable | Response | Frequency | Valid % | Mean | S.D. |
|---|---|---|---|---|---|
| 77 | Know to be true | 60 | 14.9 | | |
| | Believe to be true | 150 | 37.2 | | |
| | Do not know/unsure | 86 | 21.3 | | |
| | Believe to be false | 85 | 21.1 | | |
| | Know to be false | 22 | 5.5 | | |
| | Total | 403 | 100.0 | 2.65 | 1.13 |
| 78 | Know to be true | 11 | 2.7 | | |
| | Believe to be true | 34 | 8.4 | | |
| | Do not know/unsure | 215 | 53.3 | | |
| | Believe to be false | 114 | 28.3 | | |
| | Know to be false | 29 | 7.2 | | |
| | Total | 403 | 100.0 | 3.29 | 0.83 |
| 79 | Know to be true | 22 | 5.5 | | |
| | Believe to be true | 87 | 21.6 | | |
| | Do not know/unsure | 171 | 42.4 | | |
| | Believe to be false | 86 | 21.3 | | |
| | Know to be false | 37 | 9.2 | | |
| | Total | 403 | 100.0 | 3.07 | 1.01 |
| 80 | Know to be true | 48 | 12.0 | | |
| | Believe to be true | 126 | 31.6 | | |
| | Do not know/unsure | 147 | 36.8 | | |
| | Believe to be false | 73 | 18.3 | | |
| | Know to be false | 5 | 1.3 | | |
| | Total | 399 | 100.0 | 2.65 | 0.95 |
| 81 | Know to be true | 21 | 5.4 | | |
| | Believe to be true | 110 | 28.1 | | |
| | Do not know/unsure | 209 | 53.5 | | |
| | Believe to be false | 48 | 12.3 | | |
| | Know to be false | 3 | 0.8 | | |
| | Total | 391 | 100.0 | 2.75 | 0.77 |
| 82 | Know to be true | 15 | 3.9 | | |
| | Believe to be true | 98 | 25.2 | | |
| | Do not know/unsure | 232 | 59.6 | | |
| | Believe to be false | 39 | 10.0 | | |
| | Know to be false | 5 | 1.3 | | |
| | Total | 389 | 100.0 | 2.80 | 0.72 |

Continued

**Appendix 1: Individual Item Analysis (Continued)**

| Variable | Response | Frequency | Valid % | Mean | S.D. |
|---|---|---|---|---|---|
| 83 | Know to be true | 55 | 14.1 | | |
| | Believe to be true | 169 | 43.4 | | |
| | Do not know/unsure | 70 | 18.0 | | |
| | Believe to be false | 80 | 20.6 | | |
| | Know to be false | 15 | 3.9 | | |
| | Total | 389 | 100.0 | 2.57 | 1.08 |
| 84 | Know to be true | 34 | 8.7 | | |
| | Believe to be true | 76 | 19.5 | | |
| | Do not know/unsure | 177 | 45.5 | | |
| | Believe to be false | 73 | 18.8 | | |
| | Know to be false | 29 | 7.5 | | |
| | Total | 389 | 100.0 | 2.97 | 1.02 |
| 85 | Know to be true | 30 | 7.7 | | |
| | Believe to be true | 90 | 23.1 | | |
| | Do not know/unsure | 122 | 31.4 | | |
| | Believe to be false | 109 | 28.0 | | |
| | Know to be false | 38 | 9.8 | | |
| | Total | 389 | 100.0 | 3.09 | 1.10 |
| 86 | Know to be true | 49 | 12.6 | | |
| | Believe to be true | 121 | 31.0 | | |
| | Do not know/unsure | 65 | 16.7 | | |
| | Believe to be false | 115 | 29.5 | | |
| | Know to be false | 40 | 10.3 | | |
| | Total | 390 | 100.0 | 2.94 | 1.23 |
| 87 | Know to be true | 116 | 29.8 | | |
| | Believe to be true | 108 | 27.8 | | |
| | Do not know/unsure | 30 | 7.7 | | |
| | Believe to be false | 46 | 11.8 | | |
| | Know to be false | 89 | 22.9 | | |
| | Total | 389 | 100.0 | 2.70 | 1.56 |
| 88 | Know to be true | 57 | 14.7 | | |
| | Believe to be true | 180 | 46.3 | | |
| | Do not know/unsure | 71 | 18.3 | | |
| | Believe to be false | 64 | 16.5 | | |
| | Know to be false | 17 | 4.4 | | |
| | Total | 389 | 100.0 | 2.50 | 1.07 |

Continued

**Appendix 1: Individual Item Analysis (Continued)**

| Variable | Response | Frequency | Valid % | Mean | S.D. |
|---|---|---|---|---|---|
| 89 | Know to be true | 41 | 10.5 | | |
| | Believe to be true | 175 | 44.9 | | |
| | Do not know/unsure | 101 | 25.9 | | |
| | Believe to be false | 62 | 15.9 | | |
| | Know to be false | 11 | 2.8 | | |
| | Total | 390 | 100.0 | 2.56 | 0.97 |
| 90 | Know to be true | 53 | 13.7 | | |
| | Believe to be true | 150 | 38.7 | | |
| | Do not know/unsure | 132 | 34.0 | | |
| | Believe to be false | 47 | 12.1 | | |
| | Know to be false | 6 | 1.5 | | |
| | Total | 388 | 100.0 | 2.49 | 0.93 |
| 91 | Know to be true | 136 | 35.1 | | |
| | Believe to be true | 187 | 48.2 | | |
| | Do not know/unsure | 43 | 11.1 | | |
| | Believe to be false | 18 | 4.6 | | |
| | Know to be false | 4 | 1.0 | | |
| | Total | 388 | 100.0 | 1.88 | 0.85 |
| 92 | Know to be true | 4 | 1.0 | | |
| | Believe to be true | 14 | 3.6 | | |
| | Do not know/unsure | 68 | 17.5 | | |
| | Believe to be false | 177 | 45.6 | | |
| | Know to be false | 125 | 32.2 | | |
| | Total | 388 | 100.0 | 4.04 | 0.86 |
| 93 | Know to be true | 14 | 3.6 | | |
| | Believe to be true | 140 | 36.0 | | |
| | Do not know/unsure | 122 | 31.4 | | |
| | Believe to be false | 89 | 22.9 | | |
| | Know to be false | 24 | 6.2 | | |
| | Total | 389 | 100.0 | 2.92 | 0.99 |
| 94 | Know to be true | 20 | 5.2 | | |
| | Believe to be true | 136 | 35.1 | | |
| | Do not know/unsure | 75 | 19.4 | | |
| | Believe to be false | 124 | 32.0 | | |
| | Know to be false | 32 | 8.3 | | |
| | Total | 387 | 100.0 | 3.03 | 1.10 |

## Appendix 2: Comparison of Percentage Distribution for Common Items

| | Loerke (1993) Study | | Current Study | |
|---|---|---|---|---|
| **Response** | **Variable 43** Mean 2.35 | **Valid %** | **Variable 53** Mean 2.16 | **Valid %** |
| Know to be true | | 45.0 | | 51.2 |
| Believe to be true | | 15.0 | | 17.9 |
| Do not know | | 16.0 | | 11.3 |
| Believe to be false | | 3.0 | | 3.2 |
| Know to be false | | 20.0 | | 16.4 |
| **Response** | **Variable 44** Mean 1.72 | **Valid %** | **Variable 54** Mean 1.56 | **Valid %** |
| Know to be true | | 44.0 | | 56.2 |
| Believe to be true | | 40.0 | | 32.2 |
| Do not know | | 16.0 | | 11.1 |
| Believe to be false | | 0.0 | | 0.5 |
| Know to be false | | 0 | | 0 |
| **Response** | **Variable 45** Mean 1.96 | **Valid %** | **Variable 55** Mean 1.69 | **Valid %** |
| Know to be true | | 32.0 | | 44.7 |
| Believe to be true | | 46.0 | | 41.9 |
| Do not know | | 21.0 | | 13.4 |
| Believe to be false | | 1.0 | | 0 |
| Know to be false | | 1.0 | | 0 |
| **Response** | **Variable 46** Mean 2.00 | **Valid %** | **Variable 60** Mean 1.93 | **Valid %** |
| Know to be true | | 29.0 | | 33.9 |
| Believe to be true | | 47.0 | | 44.3 |
| Do not know | | 19.0 | | 18.0 |
| Believe to be false | | 5.0 | | 2.9 |
| Know to be false | | 0 | | 1.0 |
| **Response** | **Variable 47** Mean 2.62 | **Valid %** | **Variable 66** Mean 3.20 | **Valid %** |
| Know to be true | | 7.0 | | 4.4 |
| Believe to be true | | 26.0 | | 18.2 |
| Do not know | | 37.0 | | 41.8 |
| Believe to be false | | 22.0 | | 23.9 |
| Know to be false | | 8.0 | | 11.7 |
| **Response** | **Variable 48** Mean 3.71 | **Valid %** | **Variable 67** Mean 3.68 | **Valid %** |
| Know to be true | | 1.0 | | 1.8 |
| Believe to be true | | 6.0 | | 6.5 |
| Do not know | | 37.0 | | 33.7 |
| Believe to be false | | 33.0 | | 37.7 |
| Know to be false | | 23.0 | | 20.4 |

Continued

## Appendix 2: Percentage Distribution (Continued)

| | Loerke (1993) Study | | Current Study | |
|---|---|---|---|---|
| **Response** | **Variable 49** | **Valid %** | **Variable 68** | **Valid %** |
| | Mean 2.99 | | Mean 3.08 | |
| Know to be true | | 2.0 | | 3.0 |
| Believe to be true | | 26.0 | | 17.2 |
| Do not know | | 48.0 | | 53.6 |
| Believe to be false | | 19.0 | | 21.7 |
| Know to be false | | 5.0 | | 4.5 |
| **Response** | **Variable 50** | **Valid %** | **Variable 69** | **Valid %** |
| | Mean 3.44 | | Mean 3.35 | |
| Know to be true | | 1.0 | | 1.7 |
| Believe to be true | | 8.0 | | 8.0 |
| Do not know | | 52.0 | | 53.0 |
| Believe to be false | | 29.0 | | 27.9 |
| Know to be false | | 11.0 | | 9.5 |
| **Response** | **Variable 51** | **Valid %** | **Variable 72** | **Valid %** |
| | Mean 3.24 | | Mean 2.90 | |
| Know to be true | | 5.0 | | 7.9 |
| Believe to be true | | 17.0 | | 25.3 |
| Do not know | | 37.0 | | 41.4 |
| Believe to be false | | 26.0 | | 19.4 |
| Know-to be false | | 14.0 | | 6.0 |
| **Response** | **V 52/55** | **Valid %** | **Variable 74** | **Valid %** |
| | Mean 3.65 | | Mean 3.19 | |
| Know to be true | | 1.0/1.0 | | 2.7 |
| Believe to be true | | 4.0/3.0 | | 9.5 |
| Do not know | | 43.0/47.0 | | 58.7 |
| Believe to be false | | 35.0/32.0 | | 24.6 |
| Know to be false | | 17.0/18.0 | | 4.5 |
| **Response** | **Variable 53** | **Valid %** | **Variable 75** | **Valid %** |
| | Mean 3.43 | | Mean 3.17 | |
| Know to be true | | 1.0 | | 1.7 |
| Believe to be true | | 9.0 | | 13.2 |
| Do not know | | 48.0 | | 57.3 |
| Believe to be false | | 30.0 | | 22.3 |
| Know to be false | | 12.0 | | 5.5 |
| **Response** | **Variable 56** | **Valid %** | **Variable 76** | **Valid %** |
| | Mean 2.20 | | Mean 1.72 | |
| Know to be true | | 32.0 | | 49.8 |
| Believe to be true | | 39.0 | | 36.3 |
| Do not know | | 11.0 | | 7.0 |
| Believe to be false | | 13.0 | | 6.0 |
| Know to be false | | 5.0 | | 1.0 |

Continued

## Appendix 2: Percentage Distribution (Continued)

| | Loerke (1993) Study | | Current Study | |
|---|---|---|---|---|
| **Response** | **Variable 54** | **Valid %** | **Variable 78** | **Valid %** |
| | Mean 3.33 | | Mean 3.29 | |
| Know to be true | | 4.0 | | 2.7 |
| Believe to be true | | 11.0 | | 8.4 |
| Do not know | | 45.0 | | 53.3 |
| Believe to be false | | 28.0 | | 28.3 |
| Know to be false | | 12.0 | | 7.2 |

| **Response** | **Variable 57** | **Valid %** | **Variable 89** | **Valid %** |
|---|---|---|---|---|
| | Mean 3.47 | | Mean 2.56 | |
| Know to be true | | 3.0 | | 10.5 |
| Believe to be true | | 11.0 | | 44.9 |
| Do not know | | 40.0 | | 25.9 |
| Believe to be false | | 28.0 | | 15.9 |
| Know to be false | | 18.0 | | 2.8 |

## Appendix 3: Common Study Items

| Questionnaire Item | | Variable |
|---|---|---|
| 9) | I have read the Alberta Learning published guidelines for interpreting diploma examination results. | (V53) |
| 10) | This school's administration are aware of the published guidelines and resources available for interpreting diploma examination results. | (V54) |
| 11) | This district's administration are aware of the published guidelines and resources available for interpreting diploma examination results. | (V55) |
| 16) | Diploma examination results are used fairly at this school. | (V60) |
| 22) | Diploma examination results are used *informally* by this school's administration as part of teacher evaluation. | (V66) |
| 23) | Diploma examination results are used *formally* by this school's administration as part of teacher evaluation. | (V67) |
| 24) | Diploma examination results are used *informally* by this district's administration as part of teacher evaluation. | (V68) |
| 25) | Diploma examination results are used *formally* by this district's administration as part of teacher evaluation. | (V69) |
| 28) | Diploma examinations are used by this school's administration to determine teaching assignments. | (V72) |
| 30) | Teachers have been transferred or have asked for transfers because of poor diploma results. | (V74) |
| 31) | Teachers have been promoted to leadership roles because of good diploma results. | (V75) |
| 32) | There is a greater level of stress associated with teaching a diploma subject than a non-diploma subject. | (V76) |
| 34) | Teachers at this school have asked for non-diploma teaching assignments because of poor diploma results. | (V78) |
| 45) | At this school, students select subjects based on their anticipated success on the diploma examinations. | (V89) |

**Appendix 4: Total Variance Explained for Factors with Eigenvalues >1**

| Factor | Eigenvalue | % of Variance | Cumulative % |
|---|---|---|---|
| 1 | 7.17 | 17.06 | 17.06 |
| 2 | 3.50 | 8.33 | 25.39 |
| 3 | 2.81 | 6.70 | 32.08 |
| 4 | 2.12 | 5.05 | 37.13 |
| 5 | 1.81 | 4.32 | 41.45 |
| 6 | 1.46 | 3.47 | 44.92 |
| 7 | 1.32 | 3.14 | 48.05 |
| 8 | 1.26 | 3.00 | 51.05 |
| 9 | 1.17 | 2.78 | 53.83 |
| 10 | 1.08 | 2.57 | 56.40 |
| 11 | 1.03 | 2.46 | 58.86 |

Extraction Method: Principal Component Analysis

## Appendix 5: Study Comparison

| Study Details | Loerke (1993) Study | Current Study |
|---|---|---|
| Schools Surveyed | 42 | 47 |
| Teachers Surveyed | 805 | 452 |
| School Size Distribution* | 14 Small<br>14 Medium<br>14 Large | 19 Small<br>13 Medium<br>15 Large |
| Questionnaire Return Rate | 56.6% | 58.6% |
| Questionnaire Items | 15 | 42 |
| Biographical Questions | 8 | 10 |
| Interviews | 0 | 10 |
| Likert Scale (1-5 ) Used for Questionnaire Items | Yes | Yes |
| Quantitative Analysis: | | |
| Factor Analysis | No | Yes |
| ANOVA | Yes | Yes |
| T Tests | Yes | Yes |

* Size based on the number of diploma examinations written at the school